PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 19 July 15, 1933 Number 7

ON THE INTERPRETATION OF NOVA SPECTRA

By Donald H. Menzel and Cecilia H. Payne

Harvard College Observatory, Cambridge, Massachusetts

Read before the Academy, Monday, April 24, 1933

1.—The spectrum of a nova, before and during maximum brightness, is not unlike that of a supergiant star of Class B, A or F, with the absorption lines greatly shifted toward the violet. After maximum the spectrum changes rapidly with the development of wide, complex bright lines, undisplaced in the mean from their normal positions; complex structure, corresponding to several different velocity shifts, also appears in the absorption lines, which become progressively fainter as the bright-line spectrum increases in intensity.

It has long been realized that the violet shift of the absorption lines is a result of a rise of the absorbing material, and this upheaval of the star's surface helps to account also for the brightening of the nova, since it is evident that the rapid increase in brilliance cannot be attributed solely to increase in surface brightness; the area of the radiating surface must be increasing as well. It has been suggested that a nova outburst begins with the expansion of the star as a whole, and that near maximum brightness the star ejects a shell of nebulous matter, and then contracts and grows fainter. Though this model interprets many features of the behavior of novae, it fails in several important points, and the present note attempts to outline another model that may be more satisfactory.

2.—We suppose that the expansion is confined to the photospheric layers, which are blown away from the stellar surface at the time of the first outburst. The star then becomes a sort of planetary nebula, in structure if not in physical conditions. In the early stages the shell should still be dense enough to radiate like a stellar photosphere, receiving diluted high-temperature radiation on its inner surface, and reradiating the transmitted radiation undiluted at a lower temperature. The observed progression of the nova spectrum from an absorption spectrum to a bright-line spectrum, which soon develops the nebular lines and ultimately shows Wolf-Rayet characteristics, is in general harmony with this model; that the theory is also substantiated in considerable detail will be shown later in the present note.

3.—The profile of an absorption line in an expanding photosphere of constant (high) velocity depends upon two factors: the variation of intensity of the line over the stellar disc, and the degree of darkening of the photosphere toward the limb in the wave-lengths that border the line. In the case of a nova, where the velocity shifts are of the order of 1000 km./sec., it is unnecessary to know the form of the original line profile; we assume that the profile is rectangular and that the intensity within the line is constant. In figure 1 are shown schematically the predicted profiles corresponding to four assumptions: (*a*) absorption line of constant intensity on expanding disc of uniform brightness; (*b*) absorption line of constant intensity on expanding disc darkened toward the limb; (*c*) "chromospheric" emission present (line dark at the center of the disc, bright toward the limb); (*d*) shell effectively transparent, except perhaps to the radiation of the nuclear star. The last case, discussed by Beals,[1] gives a rectangular emission profile.

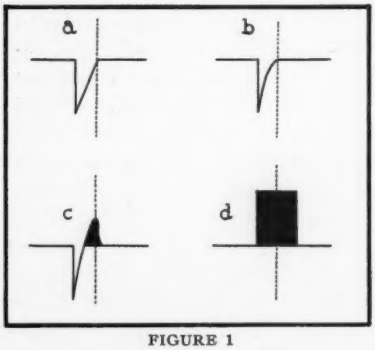


FIGURE 1

Schematic profiles of lines in the spectrum of a nova expanding with constant high velocity. Bright lines are blackened. A dotted line indicates the undisplaced wave-length; violet is to the left. The four cases represented are: (*a*) uniform disc; (*b*) disc darkened at the limb; (*c*) lines dark at the center, bright at the limb; (*d*) emission from transparent expanding shell.

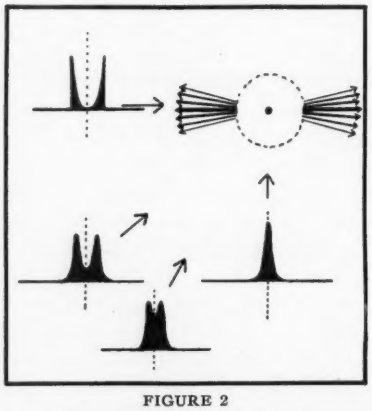


FIGURE 2

Schematic profiles given by a transparent expanding shell, ejected directionally as shown in the upper right corner. Four predicted profiles, seen as the star is viewed from the direction of the four arrows, are shown. Ejection is supposed to be more plentiful at the center of the jet. The diagram is two-dimensional; the fact that the actual case is three-dimensional does not change the effects essentially.

The foregoing assumptions are too simple. Unsymmetrical ejection of material would obviously affect the observed profile. In figure 2 is shown the predicted profile for a star with an effectively transparent radiation

shell, ejecting matter symmetrically with respect to two hemispheres, and viewed from several angles. The superposition of successive outbursts would also complicate the profile. Further, the predictions of figures 1 and 2 were made on the basis of a single constant velocity, but, as one of the writers has pointed out,[2] any symmetrical profile can be predicted by appropriate choice of velocity distribution, for an atmosphere expanding isotropically.

4.—A detailed study, now being made by the writers, of the available material of the spectra of bright novae at Harvard and elsewhere, will ultimately provide data for comparing observed and predicted profiles. Cases (*a*) and (*b*) have not at present been discriminated, but a study of good premaximum spectra should provide the necessary data. Case (*c*), "chromospheric" emission, is shown by the premaximum spectra of Nova Persei, Nova Aquilae, Nova Cygni and Nova Pictoris, and is doubtless a normal feature of the early spectra of all novae. Case (*d*), the rectangular profile, is not shown by any novae that we have at present examined. Some of the bright lines show a rounded form, and others, while sharply bounded at the red and violet, show a central dip; the two types coexist in the spectrum of Nova Aquilae, where the hydrogen lines show a saddle-shaped, and the nitrogen lines a rounded profile. Both of these types of profile could be interpreted equally well in terms of appropriate velocity distribution, or of directional ejection. The structure of individual bright lines in the spectrum of Nova Aquilae[3] clearly points to directional ejection in jets, with respect to a definite plane which may be defined by the star's equator.

5.—The interaction of the greatly widened profiles of neighboring lines is of much interest. Emission profiles overlap in a purely additive way, producing a complex structural appearance. The relation between emission and absorption lines seems to indicate that absorption takes place at the lower level of the stellar surface, while the emission lines are produced above them in the tenuous envelope. This conclusion is drawn from the fact that in the case of a close group of lines, all the emission components are present but only the violet absorption component. This observation and this interpretation also apply to the Wolf-Rayet stars.

6.—Most of the interpretation of the spectra of novae that has hitherto been made has hinged on the observations of radial velocity, and hence has dealt primarily with the state of motion of the outer shell. Extensive analysis by Wright,[4] by Spencer Jones[5] and by Stratton[6] has led to the possibility of drawing conclusions also as to the composition and constitution of the atmosphere.

7.—Recent investigations of the spectra of the W (Wolf-Rayet) stars have resulted in the almost complete identification in them of the lines of carbon, nitrogen and oxygen up to and including the three-electron stage.[7] The

temperatures of these stars, estimated by the method of Zanstra,[8] are of the order of 50,000° to 100,000°. The W stars, as has been shown by Menzel[9] and by Beals,[10] are losing their atmospheres at a slow but steady rate with high velocity, so that they are surrounding themselves with nebulous envelopes, which are however of too high a density to permit the "forbidden" nebular lines to appear. The novae present an obvious parallel, since they too are losing their atmospheres with high velocity, and since they ultimately show a characteristic W spectrum.

8.—When a developing nova is regarded from this standpoint the successive spectral changes are seen to form a logical sequence. The absorption stage is followed by the successive appearance of normal bright lines and forbidden bright lines. The absorption spectrum changes, as the star passes maximum, in a way that indicates a rapid increase of excitation; for example, the appearance in Nova Aquilae[11] of the lines of N V indicates that the star has risen from perhaps 10,000° to an ionization temperature of about 100,000° in the course of a few months. The normal bright lines appear in order of excitation energy, the bright-line spectrum lagging appreciably behind the absorption spectrum of similar excitation. The forbidden bright lines appear also in order of excitation energy. The observed forbidden transitions usually occur for atoms whose ground states have the configurations p^2, p^3 and p^4. These configurations all produce two metastable terms, with three possible transitions between each other and the ground state, giving rise to what we have termed nebular, auroral and transauroral lines. For any one atom the auroral lines appear first, and are followed after a few days (Nova Aquilae) or a few weeks (Nova Pictoris) by the nebular lines. So regular is the sequence of appearance of lines of a given type that the date at which a line first appears furnishes an effective means of line identification in the nova spectrum. Reference should be made to an earlier discussion of forbidden lines in astrophysical sources.[12]

The nebular lines gradually increase in strength, and, in some novae, eventually surpass the auroral lines. They are almost always more prominent in the spectra of nebulae than the auroral lines. Bowen[13] has suggested that the difference of intensity of the nebular and auroral lines of O I in the spectra of the terrestrial aurora and the planetary nebulae may be a result of different relative lives for the metastable states.

9.—The analysis of the spectrum of the developing nova along these lines promises to be an extensive investigation. As a general illustration of the data, tables 1, 2 and 3 present the observations of prominent lines for three typical novae. The tables list the atoms whose spectra appear on the various dates, with the excitation potentials of their strongest lines. Since the material of the nova is probably not in thermodynamic equilibrium, the significant excitation potential is at present uncertain.

For convenience we have listed the cumulative potentials (in volts) counted from the ground state of the neutral atom up to the state of the atom from which the line in question arises.

Any astrophysical study that employs marginal appearances is open to question, since the dates of the appearances are, no doubt, influenced by factors other than the excitation potential, for example, by abundance. The dates upon which the various lines were at maximum appear to be much more valuable indices of the excitation energies, a conclusion supported by the meager evidence now available. The authors plan to investigate this point further.

TABLE 1

SPECTRAL DEVELOPMENT OF NOVA PERSEI

DATE 1901	ABSORPTION	EMISSION			
		PERMITTED	AURORAL	NEBULAR	TRANSAURORAL
Feb. 22	Mg I 2.7				
	Ca II 6.1				
	H 10.2	H 12.0			
	Mg II 16.4	Mg II 19.2			
	Si II 17.9	Si II 20.9			
	He I 20.9				
	Al II 21.0				
	C II 29.2				
	N II 32.9				
	O II 39.2				
	Al III 40.3				
	He II 72.6				
Feb. 23		Sr II 8.6			
		Ti II 10.6			
Feb. 24		Mg I 5.1			
		Ca II 9.2			
		Fe II 13.2			
Feb. 25		He I 23.0	O I 4.2	N I 2.4	S I 2.3
					C I 2.7
Feb. 28			?N II 18.5	?O II 16.9	
Mar. 2	N II*				S II 13.3
Mar. 19		He II 75.3			
		N III 77.0			
Apr. 12			O III 53.8		
Apr. 28†			Ne IV 132	O III 51.0	
				Ne III 65.4	

* Reappearance.

† No observations since April 13. These lines may have appeared earlier than April 28.

10.—The excitation analysis of the bright-line spectra of novae furnishes a basis of comparison for novae that developed at very different rates. There are of course differences of detail, but all novae at the same excitation stage show approximately similar spectra.

TABLE 2

SPECTRAL DEVELOPMENT OF NOVA AQUILAE

DATE 1918	ABSORPTION	EMISSION			
		PERMITTED	AURORAL	NEBULAR	TRANSAURORAL
June 8	H 10.2	H 12.0			
	Mg II 16.4				
	Si II 17.9				
June 9	Fe I 0.0				
	Ti II 7.9				
	Fe II 10.7				
June 10		Fe II 13.2			
June 12		He I 23.0			
June 14	N III 74.3				
June 15	(before)		O I 4.2		S I 2.3
June 15		He II 75.3			
June 16		N III 77.0			
June 18			N II 18.5	N II 16.4	
				O II 16.9	
June 24	N IV 141.5				
July 3		N IV 145		Ne III 65.4	
July 5			O III 53.8		
July 8				O III 51.0	
July 12			Ne IV 132		
Aug.	N V 225				

The most obvious difference between the bright-line spectra of different novae is found in the behavior of the lines of ionized iron. In only two novae (Nova Pictoris[14] and Nova Aurigae[15]) have the forbidden Fe II lines appeared during the outburst, though the celebrated ex-nova η Carinae has developed them with great prominence[16] since 1889, and shows them still. Evidently the novae that develop the most slowly exhibit the forbidden Fe II spectrum, probably because only for such objects is the requisite low density attained before the temperature rises so high that all the Fe II is further ionized. The slow fall in brightness and the slow spectral changes of Nova Pictoris and Nova Aurigae are almost certainly associated with a slower rise of temperature than takes place for most novae. Possibly some of the fast developing novae may, in a similar way, fail to exhibit other forbidden lines of low excitation; further observations are required on this point.

11.—In the final stages of the nova the forbidden lines usually disappear. This may be attributed in part to the general increase in ionization and to the decreased brightness of the star, but the persistence of the lines of H and He II shows that an abundant supply of energy is still available. Bowen[17] and Zanstra[18] look with favor on the hypothesis that the forbidden lines in nebulae arise from collisional excitation rather than from recombination. In the late stages of a nova collisions would then be too infrequent, owing to the low density, to excite an appreciable number of atoms.

TABLE 3

SPECTRAL DEVELOPMENT OF NOVA PICTORIS

DATE 1925	ABSORPTION	EMISSION			
		PERMITTED	AURORAL	NEBULAR	TRANSAURORAL
May 25	Al I 0.0				
	Ca I 0.0				
	Fe I 0.0				
	Ti I 0.8				
	Mg I 2.7				
	Sr II 5.6				
	Ca II 6.1				
	Ti II 8.0				
	Zr II 8.1				
	H 10.2				
	Fe II 10.6				
	Mg II 16.4				
June 10		Ca I 2.9			C I 2.7
		Mn I 3.1			
		Fe I 4.5			
		Ca II 9.2			
		Sc II 9.8			
		Ti II 10.6			
		H 12.0			
		Fe II 13.2			
June 12		Na I* 2.1		N I* 2.4	S I* 2.3
		Y II 9.7			
		Zr II 10.5			
		Cr II 13.5			
		He I* 23.0			
June 25					?Si I 1.9
Aug. 17			Fe II 10.7	Fe II 11.0	
Aug. 24					S II 13.3
Dec. 21			O III 53.8		
1926					
Jan. 7		He I† 23.0		O III 51.0	
				Ne III 65.4	
Mar. 15		He II‡ 75.3	Ne IV 132		
		N III‡ 77.0			
Sept. 7			F III§ 56		

* Earliest record.

† Reappearance.

‡ These lines probably appeared much earlier; this is the first date on which they are sharp and well resolved.

§ This is the date given by H. Spencer Jones. The Harvard plates seem to be systematically earlier than the Johannesburg plates and this line may have appeared before September 7.

The mechanism of excitation of the bright lines will be discussed in a later paper.

The bright-line spectra of novae can be interpreted in a general way, in terms chiefly of rapidly rising temperature, rapidly falling pressure and

rapidly increasing dilution of radiation. But the problem is made somewhat more complex by the probable existence of differences of chemical composition. In comparing various novae, the irregularity in the distribution of carbon, noted for the W stars,[19] also appears. In the development of some novae the permitted lines of several successive stages of the carbon atom have been noted; from others they are absent. Unfortunately there are no accessible forbidden lines of these stages of the carbon atom. The apparent high abundance of neon and fluorine in the atmospheres of the novae leads one to suppose that these atoms are cosmically very abundant; it should be noted that neon is conspicuous in the nebular spectra, while fluorine is not. A search for the stellar absorption lines of these elements is to be desired; and a search for irregularities in cosmic abundance is now in progress.

We wish to express our appreciation to Dr. J. H. Moore and Dr. W. H. Wright of the Lick Observatory for valuable suggestions made while this paper was in manuscript.

[1] C. S. Beals, *Mon. Not.*, **91**, 966 (1931).
[2] C. H. Payne, *Harv. Bul.*, 874 (1930).
[3] J. H. Moore and C. D. Shane, *Lick Obs. Bul.*, No. 322 (1919).
[4] W. H. Wright, *Publ. Lick Obs.*, **14** (1924).
[5] H. Spencer Jones, *Ann. Cape Obs.*, **10** (1931).
[6] F. J. M. Stratton, *Ann. Sol. Phys. Obs.*, Cambridge, **4**, Part 1 (1920).
[7] C. H. Payne, *Zeits. f. Astrophys.*, in press.
[8] H. Zanstra, *Astrophys. Journ.*, **65**, 50 (1927).
[9] D. H. Menzel, *Publ. Astr. Soc. Pac.*, **41**, 344 (1929).
[10] C. S. Beals, *Mon. Not.*, **90**, 202 (1929).
[11] F. J. M. Stratton, *Obs.*, **55**, 207 (1932).
[12] J. C. Boyce, D. H. Menzel and C. H. Payne, these PROCEEDINGS, **19**, 581–591 (1933).
[13] I. S. Bowen, *Phys. Rev.*, **36**, 600 (1930).
[14] H. Spencer Jones, *Ann. Cape Obs.*, **10**, 108 (1931).
[15] D. Gill, *Roy. Soc. Proc.*, **68A**, 456 (1901); H. Spencer Jones, *Ann. Cape Obs.*, **10**, 128 (1931).
[16] P. W. Merrill, *Contr. Mt. W. Obs.*, 354 (1928).
[17] I. S. Bowen, *Astrophys. Journ.*, **67**, 1 (1928).
[18] H. Zanstra, *Dom. Astrophys. Obs. Publ.*, **4**, 209 (1931).
[19] C. H. Payne, *Zeit. f. Astrophys.*, in press.

CHROMOSOME NUMBER AND MORPHOLOGY IN NICOTIANA VI. CHROMOSOME NUMBERS OF FORTY SPECIES[1]

BY THOMAS HARPER GOODSPEED

DEPARTMENT OF BOTANY, UNIVERSITY OF CALIFORNIA

Communicated May 29, 1933

In the University of California Botanical Garden over 60 species and varieties of the genus *Nicotiana* have been under investigation, together with species of such related genera as *Petunia*, *Fabiana*, *Salpiglossis* and *Nierembergia*. Morphological, distributional and cytogenetic evidence indicates the desirability of recognizing among the races of *Nicotiana* under observation, the following 40 species. The list is arranged in order of date of description.

glutinosa L.
paniculata L.
rustica L.
tabacum L.
tomentosa Ruiz & Pavon
undulata Ruiz & Pavon
longiflora Cav.
plumbaginifolia Viv.
bonariensis Lehm.
repanda Willd.
suaveolens Lehm.
Langsdorffii Weinm.
alata Link & Otto
glauca Grah.
acuminata (Grah.) Hook.
solanifolia Walp.
rotundifolia Lindl.
corymbosa Remy
Miersii Remy
Cavanillesii Dun.
trigonophylla Dun.
cordifolia Phil.
wigandioides Koch. & Fint.
megalosiphon van Heurck & Müll.
attenuata Torr.
Bigelovii S. Wats.
nudicaulis S. Wats.
caudigera Phil.
pampasana O. Kuntze
sylvestris Speg. & Comes
Stocktoni Brandg.
Sanderae hort.
Debneyi Domin
Raimondii Macbride
nesophila Johnston
tomentosiformis Goodspeed
4 unnamed Australian species

The haploid chromosome numbers characteristic of these 40 species are 9, 10, 12, 16, 20, 22, 24 and 32, and in the following table the species are arranged under these various chromosome number classes.

For the sake of completeness, chromosome numbers of species earlier published by the author and by others are included in table 1 along with those first reported in the above list. Our representatives of the species listed have received detailed taxonomic study to assure nomenclatorial accuracy. For reasons to be discussed elsewhere, it appears desirable to distinguish *plumbaginifolia* from *longiflora*, *caudigera* from *acuminata*, *Raimondii* from *glauca* and *nesophila* from *Stocktoni*. The status of *wigandioides* and *tomentosiformis* has already been commented upon (Goodspeed[2,3]).

As indicated in footnote 2 to table 1, separate discussion of the *Nicotiana* species peculiar to Australia proper, Lord Howe Island and Rarotonga will be forthcoming. It should be noted, however, that decision to recog-

TABLE 1

CHROMOSOME NUMBERS OF *Nicotiana* SPECIES[1]

9	10	12	16	20	22	24	32
alata bonariensis Langsdorffii Sanderae	longiflora plumbaginifolia	acuminata attenuata caudigera Cavanillesii cordifolia corymbosa glauca glutinosa Miersii pampasana paniculata Raimondii solanifolia sylvestris tomentosa tomentosiformis trigonophylla undulata wigandioides	suaveolens and 3 unnamed species[2]	megalosiphon[2]	rotundifolia[2]	Bigelovii Debneyi[2] nesophila nudicaulis repanda rustica Stocktoni tabacum	unnamed species[2]
4	2	19	4	1	1	8	1

[1] Expressed as number of pairs of chromosomes and based upon counts in P. M. C. and in root-tips.
[2] Australasian species concerning the cytotaxonomy of which an article by Miss Helen-Mar Wheeler is in preparation.

nize as new species 4 races allied to *suaveolens* and obtained from the wild state came only after thorough comparative study of all available herbarium material of types. Of special interest are two 32_{II} races, one of which represents an instance of polyploidy within the species, being a replica of typical *suaveolens* on a large scale, while the other is apparently a distinct species.

With the exception of *N. sylvestris* and *paniculata*, all the species listed above give evidence, from herbarium specimens or in the living state, of being more or less polymorphic. As a result, named varieties of many of them are recognized (cf. Millán[4]), but often, particularly in botanical gardens, varieties, and in some cases minor ones, are given species designations. It might be noted in this connection that a considerable number of varieties of *tabacum*, *rustica*, *Bigelovii*, *acuminata*, *longiflora* and *Cavanillesii* examined show no variation in chromosome number.

Other authorities for chromosome number determinations are cited by Tischler[5] and Gaiser.[6] The only species not in the above list (table 1) mentioned by them are *cerinthoides*, *noctiflora*, *petiolaris*, *Forgetiana*, *viscosa*, *Palmeri*, *angustifolia*, *Rusbyi* and *clarionensis*. The species referred to by others as *Rusbyi* is listed here as *tomentosiformis* (Goodspeed[2,3]). The designation *clarionensis* was never published (cf. Johnston[7]), this species being referred to here as *Stocktoni*. De Vilmorin and Simonet[8] reported *cerinthoides*, *Forgetiana* and *noctiflora* as possessing 9_{II} and *trigonophylla*, *solanifolia* and *petiolaris* as 24_{II}. Seed of what was grown by them under these designations (except *petiolaris*) was kindly supplied by M. de Vilmorin and when grown here gave the following results: "*noctiflora*" = typical, blue-pollened *Langsdorffii*; "*Forgetiana*" and "*cerinthoides*" = apparently derivatives of *Langsdorffii* × *alata*; "*solanifolia*" and "*trigonophylla*" = *rustica* varieties. *N. petiolaris* is a designation

usually applied to certain *rustica* or *tabacum* varieties. Gaiser[6] lists "*angustifolia*" ($n = 10$) as reported by Clausen.[9] Although the taxonomic situation in regard to this species designation is confused, it is clear that the race referred to by Clausen is a *longiflora* variety. The following species for which Christoff[10] gives chromosome number determinations are also listed by Gaiser: "*viscosa* (?)" ($n = 24$), "*Palmeri* (?)" ($n = 12$), and "*solanifolia* Wolf—*cardiophylla* Ph."* ($n = 12$). There is no evidence as to the identity of Christoff's "*viscosa*" since if it had been an *attenuata* variety, as he suggests, the chromosome number should have been $n = 12$ rather than $n = 24$. Christoff refers "*Palmeri*" to *trigonophylla*.

Some 50 interspecific hybrids in *Nicotiana* have been obtained here and have included three amphdiploid types. To *digluta* ($n = 36$), an amphidiploid *glutinosa* × *tabacum* hybrid (cf. Clausen and Goodspeed[11]), can now be added the amphidiploid *glutinosa* × *tomentosa* hybrid, *diglutosa* ($n = 24$), and *disualovii* ($n = 40$), an amphidiploid *suaveolens* × *Bigelovii* hybrid. If the chromosome numbers of these three "species" are added to the series given in table 1, the complete list is 9, 10, 12, 16, 20, 22, 24, 32, 36 and 40 pairs.

Among by-products of recent cytogenetic investigations in *Nicotiana* carried on here are numerous aneuploid types in various species—*alata*, *Langsdorffii*, *sylvestris*, *paniculata*, *suaveolens*, *tabacum* and *rustica*. Tetraploid *Langsdorffii*, *sylvestris* and *tabacum* have been studied. Haplonts within all chromosome number classes doubtless occur; those at present known are in the 9_{II}, 12_{II} and 24_{II} groups, i.e., *glutinosa*, *tabacum*, *nudicaulis;* and *Langsdorffii* (Kostoff[12]). A variety of quantitative chromosome alterations have been established in stable derivatives of *tabacum* and *sylvestris* as products of treatment with high frequency radiation. A number of sesquidiploid forms have also been obtained (cf. Webber[13]).

The chromosome numbers of related genera have in a few cases been studied. *Petunia* species have for some time been known to possess 7_{II} and more recently a 9_{II} species has been reported by Ferguson and Cooldige.[14] This last is of obvious interest since a group of *Nicotiana* species shows this number and because hybrids between members of it and *Petunia* have been reported (cf., however, Kostoff[15]). The count of 22_{II} for *Salpiglossis sinuata* reported by de Vilmorin and Simonet[8] can be confirmed. For *Fabiana imbricata* and *Nierembergia frutescens*, the numbers, not before reported, are in each case $2n = 18$.

Evidence which has for some years been accumulating will make possible

* Concerning Christoff's "*solanifolia* Wolf—*cardiophylla* Ph." there is only the matter of the authority for *solanifolia* to be noted. He apparently misinterpreted "Wolf" for "Walp." on the herbarium sheet from which the seed was obtained, and this mistake is referred to because it has been perpetuated by Gaiser[6] where, in addition, the authority for *cardiophylla* is stated to be "Rh." instead of Ph.

a partially complete description of species origins and relationships in the genus *Nicotiana*, based upon morphological, cytogenetic and other studies. A few postulates which appear justified and are pertinent in the present connection may be mentioned: 1. Certain genetic groups characterized by similarity both in chromosome number and in certain basic morphological features have been recognized. The "alata group" includes *alata*, *Langsdorffii*, *Sanderae* and *bonariensis* ($n = 9$); *longiflora* and *plumbaginifolia* ($n = 10$); the "tomentosa group" ($n = 12$) consists of *tomentosa*, *tomentosiformis*, *glutinosa* and *wigandioides*; the "glauca group" ($n = 12$) of *glauca*, *Raimondii*, *paniculata*, *solanifolia*, *cordifolia* and *undulata*; the "corymbosa group" ($n = 12$) of *corymbosa* and *Miersii*; the "Cavanillesii group" ($n = 12$) of *Cavanillesii* and *pampasana*; the "acuminata group" ($n = 12$) of *acuminata*, *caudigera* and *attenuata*; the "repanda group" ($n = 24$) of *repanda*, *Stocktoni* and *nesophila*. The "suaveolens group," consisting of *suaveolens*, *Debneyi*, *rotundifolia* and *megalosiphon*, is a natural one despite the distinctions in chromosome number involved. 2. Certain, if not all, of the 24_{II} species are of allopolyploid origin. Thus *tabacum* ($n = 24$) may be traced to derivatives of hybrids between the "tomentosa group" ($n = 12$) and *sylvestris* ($n = 12$); *rustica* to derivatives of hybrids within the "paniculata group" ($n = 12$), *Bigelovii* obviously involves in its origin *attenuata* ($n = 12$) and probably *trigonophylla* ($n = 12$), while the origin of the "repanda group" and *nudicaulis* is not so clear, due perhaps to the absence of living descendants of their progenitors. The occurrence in our cultures of an amphidiploid between two 12_{II} species—*diglutosa*—lends some weight to the postulated amphidiploid origins just mentioned. That higher chromosome races of *Nicotiana* can be derived is indicated by our *digluta* and *disualovii* and by the results of Kostoff,[15] Lammerts[16–18] and Rybin.[19]

[1] The investigations reported upon were supported by a grant from the Board of Research, University of California.

[2] Goodspeed, T. H., *Bot. Gaz.*, **93**, 340–341 (1932).

[3] Goodspeed, T. H., *Ostenia*, Montevideo, 309–314 (1933).

[4] Millán, A. R., *Rev. Fac. Agr. y Vet.*, **6**, 169–216 (1928).

[5] Tischler, G., *Tabulae Biologicae*, **4**, 1–83 (1931).

[6] Gaiser, L. O., *Bibliogr. Genetica*, **6**, 171–466 (1930).

[7] Johnston, I. M., *Proc. Calif. Acad. Sci.*, Ser. 4, **20**, 9–104 (1931).

[8] Vilmorin, R. de, and Simonet, M., *Compt. Rend. Acad. Sci.*, Paris, **184**, 164–166 (1927).

[9] Clausen, R. E., *Verh. V. Intern. Kongr. Vererbwiss., Zeitschr. indukt. Abst. u. Vererb.*, Supplb., **1**, 547–553 (1928).

[10] Christoff, M., *Genetics*, **13**, 233–277 (1928).

[11] Clausen, R. E., and Goodspeed, T. H., *Genetics*, **10**, 279–284 (1925).

[12] Kostoff, D., *Zeitschr. Zellforsch. u. mikr. Anat.*, **9**, 640–642 (1929).

[13] Webber, J. M., *Univ. Calif. Publ. Bot.*, **11**, 319–354 (1930).

[14] Ferguson, M. C., and Coolidge, E. B., *Am. J. Bot.*, **19**, 644–658 (1932).

[15] Kostoff, D., *Genetica*, **12**, 33–139 (1930).

[16] Lammerts, W. E., *Genetics*, **16**, 191–211 (1931).
[17] Lammerts, W. E., *Cytologia*, **4**, 38–45 (1932).
[18] Lammerts, W. E., *Ibid.*, **4**, 46–51 (1932).
[19] Rybin, W., *Ber. d. deutsch. bot. Ges.*, **47**, 385–394 (1929).

---

*YOUNG'S MODULUS AND POISSON'S RATIO WITH REFERENCE TO GEOPHYSICAL APPLICATIONS**

BY W. A. ZISMAN

DEPARTMENT OF GEOLOGY AND GEOGRAPHY, HARVARD UNIVERSITY

Communicated May 20, 1933

Scope of the Investigation.—The theory of elasticity furnishes the following relations among the constants describing the elastic behavior of an isotropic and homogeneous medium obeying Hooke's law:

$$V^2 = \frac{E}{\rho} \frac{(1-\sigma)}{(1+\sigma)(1-2\sigma)},$$

and

$$\left(\frac{V}{v}\right)^2 = \frac{2(1-\sigma)}{1-2\sigma},$$

where V is the velocity of a longitudinal elastic wave in an infinite volume of the material, v is the velocity of a transverse wave, E is Young's modulus, σ is Poisson's ratio and ρ is the density of the medium. In rocks, V and v can be determined by seismological methods. Assuming perfect elasticity and isotropy for rocks, E and σ can be calculated. Conversely, on the same assumption one can predict V and v, if he can measure E and σ by statical methods in the laboratory, at the range of stresses involved in the propagation of earthquake waves.

The validity of the underlying assumption can be tested. With the explosion-wave method V and v are measured in an exposed terrane and thence the elastic constants are calculated. These are then compared with the constants measured statically on specimens of rock from the same terrane and at the corresponding extremely low range of stress-differences. If the respective values do not agree, the rock is not ideally elastic or homogeneous, and corrections must be applied before the constants determined from V and v, measured seismologically, can be used for identifying the rock where hidden underground. In practice, the elastic coefficient useful in such diagnoses is cubic compressibility. Hence the geophysicist needs data concerning the statical compressibility of an extensive series of standard rocks at varying hydrostatic pressures and under the

minute additional stress corresponding to that involved in wave propagation. Here published data are incomplete and furnish a poor foundation for geophysical research, especially for the outer layers of the earth.

The wide range of published values of E and σ for rocks may be illustrated by the case of granite.

The Annual Report of the Watertown Arsenal for 1895 gives values of E for this rock with limits between 3×10^{11} and 6.5×10^{11} dynes per square centimeter; and of σ with limits between 0.07 and 0.25. These represent statical tests of American granites when subjected to unilateral pressure in air.

Adams and Coker[1] found for American granites E ranging from 4.0×10^{11} to 6.3×10^{11}, with σ between 0.20 and 0.26. On page 9 of their report, they mention Nagaoka's values of E for granites, ranging from 1.1×10^{11} to 4.2×10^{11}. Adams and Coker, like the experimenter at the Watertown Arsenal, used a statical testing machine, whereas Nagaoka worked with the flexure method at loads relatively and absolutely small.

Again, Bach[2] states that E for German granites is 2.4×10^{11}.

Similar discrepancies are found in the case of practically every other type of rock investigated, and are clearly explicable not only by real differences in the specimens of each of the rock species, but also by differences in the methods of measurement. Quantitative study of both of these controls over results for E and σ, as well as new measurements of cubic compressibility, were suggested by the Harvard Committee on Geophysics. The conclusions from this study, so far as they relate to E and σ, are described in the present paper.

Apparatus.—The study of the elastic behavior of most rocks is rendered difficult by the low stresses necessary in order to avoid departure from Hooke's law. It is therefore essential to use apparatus of the greatest possible sensitivity for the measurement of the strains. Instruments were consequently built to measure E and σ by a compression test in which the applied stress rarely exceeded one thousand pounds per square inch. This meant that, if the measured values of the elastic constants were to be accurate to one per cent, the instruments must be capable of detecting strains of the order of magnitude of one part in ten million.

The heterogeneity and coarseness of texture expected in most rocks made it wisest to employ samples of a size sufficient to guarantee results fairly typical of the respective terranes. The specimens were rock cores two inches in diameter and about ten inches in length. Each rock was ground with emery in order to form a good right cylinder. The ends of the cylinder were compressed in the jaws of a testing machine and the change in diameter and change of length between two points on the cylinder wall were measured as a function of the load. From the changes in length and diameter E and σ are directly calculated. Cylinders were

cored from large blocks in directions mutually perpendicular to one another, permitting a study of the variation of the elastic properties with direction. The expression "hardest splitting way" used in table 4 means that the axis of the cylinder is in the plane of most difficult cleavage.

The gauge constructed to measure E was a modified form of the well-known Bauschinger rocking-mirror strain apparatus. It measured the change in length of the rock between two points twelve centimeters apart. The sensitivity attained was such that vibrations in the laboratory often caused great trouble. Hence many of the experiments were made at night. With a telescope magnifying thirty-five times to observe the mirror deflections, an illuminated scale at a distance of 4.5 meters from the mirror and a rocking knife-edge but one millimeter wide, a change of length of the specimen of 10^{-7} cm. caused a millimeter deflection in the telescope.

The reliability of this strain gauge was proved by comparing the result for a cylinder of soft steel with the value of E obtained by Professor G. W. Pierce on the same cylinder by a magnetostriction method of great accuracy. The two measurements agreed within one per cent. Another good check was obtained with a cylinder of stainless steel. The observed E of soft steel was 21.7×10^{11} and of stainless steel 22.1×10^{11} dynes/cm.^{2}. It is a pleasure to thank Professor Pierce for his measurements.

The apparatus for the measurement of σ is an improved form of a well-known type, utilizing a rocking mirror and a mechanical lever magnification. A detailed description will be found in the "Review of Scientific Instruments," June, 1933.

The improvements incorporated in this gauge make it capable of measuring σ to one per cent, and yet the instrument is rugged and easily manipulated. The sensitivity was such that with the same telescope and scale-to-mirror distance used for the Young's modulus experiment it was possible to obtain a scale deflection of 1 mm. for an increase of 10^{-7} cm. in the diameter of the specimen.

The behavior of the Poisson's ratio gauge was checked by measuring the lateral strains in the same specimens of steel used to check the Young's modulus apparatus. In three successive and independent runs on soft steel the values of 0.296, 0.303 and 0.304 were obtained. The last check was made four months after the first two, during which time most of the rocks were studied. Another check was obtained by computing the cubic compressibility, β, of the steel, using the measured values of E and σ and the relation from the theory of isotropic elastic bodies

$$\beta = \frac{3}{E}(1 - 2\sigma).$$

The result agreed within five per cent with P. W. Bridgman's published

measurement of the cubic compressibility of soft steel under hydrostatic pressure. The agreement is fair, especially in view of the uncertainty that the rolled steel cylinder was truly isotropic.

The compression on the specimen ends was obtained with a new Baldwin-Southwark "Emery" testing machine of sixty thousand pounds maximum loading capacity. This was placed at our service through the courtesy of Professor A. Sauveur of the Harvard Department of Metallurgy.

The rock densities given in table 4 were obtained from the dimensions and mass of the rock cylinders.

Measurement of Young's Modulus (E).—Elastic hysteresis. If the strain is plotted along one axis of coördinates and the load is plotted along the other, the curve for decreasing loads never coincides with that for increasing loads. A typical hysteresis loop for surface Quincy granite, specimen No. 1, is shown in figure 1. If a succession of such hysteresis loops is made on the same specimen, it is found that the width of the loops decreases rapidly in the first two cycles and then becomes fairly constant. The method of obtaining Young's modulus from such a graph was as follows: To find E for a mean load of 500 pounds, the strain was measured successively for loads of 0, 250, 500, 750, 1000, 900, 750, 500, 250 and 0 pounds. This cycle of loads was carried through about five times in the space of two hours. The graph for any one of the last three runs may be represented by figure 1. In general, if the value of E was desired, corresponding to a hysteresis cycle for which the maximum load was F_2 and the minimum load was F_1, the following procedure was adopted (see Fig. 2). A series of five or six hysteresis cycles was made, each requiring ten to fifteen minutes to complete. For each cycle the load was adjusted in steps commencing at F_1, increasing in four to six steps to a maximum load F_2, and decreasing in four to six steps until F_1 was reached again.

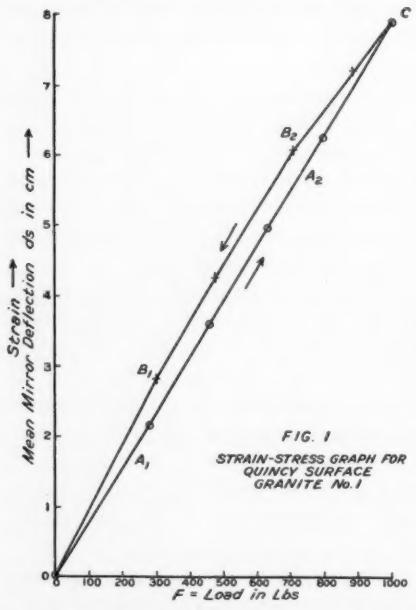


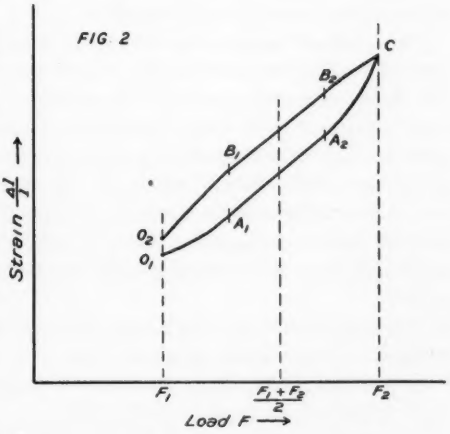
FIG. 1

STRAIN-STRESS GRAPH FOR QUINCY SURFACE GRANITE No. 1

Permanent set was shown by the lack of coincidence of points O_1 and O_2. Elastic after-effect was shown by a time change of the position of point O_2. It was always true that O_2 approached closer to O_1 with passage of time, although O_2 rarely settled down to coincidence with O_1.

It was always found true that at least half the paths of the curves, say, from A_1 to A_2 and B_1 to B_2, were practically straight lines which differed in slope by from one to five per cent. The mean slope of the lines A_1A_2 and B_1B_2 was used to calculate Young's modulus. This value of E was then identified in this report as that corresponding to the mean load $\frac{1}{2}(F_1 + F_2)$.

Thus for the cycle shown in figure 1, with the load F_1 at zero, and the loads at A_1, A_2 and C, respectively, 250, 750 and 1000 pounds, the mean of the slopes of A_1A_2 and B_1B_2 was used to compute the value of E at the load $\frac{1}{2}(F_1 + F_2) = 500$ pounds. This corresponds to a mean load of 11.2 kg./cm.^{2}.



The value of E obtained in this manner was found to be the same for all hysteresis loops after the second.

Variation of E with Mean Stress.—It was found that E varies greatly with the mean stress used. Thus the following data were obtained for Quincy surface granite, cylinder No. 2:

TABLE 1

F_1, POUNDS	F_2, POUNDS	MEAN STRESS, KG./CM.^{2}	E, DYNES/CM.^{2}
0	500	2.8	2.39×10^{11}
0	1000	11.2	2.58
500	1500	22.4	3.07
1000	2000	33.6	3.64
2000	3000	56.0	4.05
3000	4000	78.4	4.73

The results for a Rockport granite specimen with axis in plane of "easiest splitting" were:

TABLE 2

MEAN STRESS, KG./CM.^{2}	E, DYNES/CM.^{2}
0	3.48×10^{11}
11.2	3.95
22.4	4.10
44.8	4.35
67.2	4.60
89.6	4.85
112	5.10
132	5.35
157	5.60
179	5.85
202	6.10

In table 2 the value of E for zero load was obtained by extrapolating on a graph of mean load against E.

The great variation of E with different loads is the cause of much of the apparent disagreement among the results of previous experimenters. The high values obtained by Adams and Coker are due to the fact that they measured E by taking hysteresis loops for a range of loads from a minimum of 1000 pounds per square inch to a maximum of 9000 pounds per square inch. When their result for granite is compared with the one here reported for a mean stress of 4500 pounds per square inch, the agreement is quite satisfactory. The low values of Nagaoka, mentioned on a preceding page, are at least partly due to his use of much smaller mean stresses.

The final results for E at two values of the mean stress are given in table 4.

It was found that those rocks which showed the smallest change of E with stress had the least porosity.

The range in values of E for the rocks investigated is considerable—from about 1.5×10^{11} for marble to 10.0×10^{11} for trap rock. The rocks with the largest values of E show the smallest variation of E with load. This seems related to the porosity. Rocks free from cracks and pores should exhibit little change of E with stress. One would expect rock at great depths in the earth to be in such a state We cannot regard the value of E, derived on a surface specimen for a mean stress of 100 kg./cm.^{2} with this static apparatus, as equal to that for the same type of rock at a depth in the earth where the rock pressure is 100 kg./cm.^{2}, for at that depth the porosity may be much less. Again, the mean stress of 100 kg./cm.^{2} used in the testing machine is a stress applied only over the ends of the rock cylinder, whereas in the earth at great depths the pressure on the rock is more nearly hydrostatic. If such loads as are available with the testing machine could completely close the pores and cracks in the rocks, we might then obtain a limit ng value of E as the mean

stress used is increased in steps. The limiting value might then be identified as that existing at depths in the earth great enough to insure true compactness of the rocks. Actually, all those tested, except Vinal Haven trap, still showed increase of E up to the highest mean stresses applied (3000 pounds per square inch).

The data of Adams and Coker and of others should therefore be used only for the range of stresses that they employed.

Anisotropy of Rocks.—The variation of E with direction in the rocks was large in some cases. This was especially true of Vermont marble, layered gneiss and limestone. On the other hand, we find in Vinal Haven trap and in Quincy granite from a depth of 235 feet almost perfect isotropy. It should not be concluded, however, that only layered or bedded rocks show great anisotropy, for French Creek norite is decidedly anisotropic.

Measurements of Poisson's Ratio.—Elastic hysteresis. There was much less hysteresis in the stress-strain graphs for the lateral strains as compared with that obtained for the longitudinal strains. Often there was but a trace of hysteresis and evidence of elastic after-effect or permanent set was lacking. This difference in behavior is, for the most part, due to the fact that E was measured first, and afterward the same rock cylinders were used to measure σ. Obviously, then, the rock had been thoroughly pressure-seasoned before the commencement of the determination of σ.

In measuring σ, the same method as that for Young's modulus, namely, repeating hysteresis cycles and obtaining the mean slope of the stress-strain cycle, was employed.

Variation of Poisson's Ratio with Mean Stress.—It was found that Poisson's ratio also varies greatly with the mean stress applied. Table 3 gives the results for the cylinder of Rockport granite whose values of E as a function of mean stress are listed in table 2.

TABLE 3

MEAN LOAD IN POUNDS	POISSON'S RATIO
0	0.080
1000	0.088
2000	0.097
3000	0.105
4000	0.114
5000	0.122
6000	0.131
7000	0.141
8000	0.151
9000	0.162

Here the value of Poisson's ratio for zero load was obtained by extrapolation on a graph of σ as a function of mean load.

This rapid variation of σ with mean load clearly explains the differences in the results of various experimenters. Those using large mean stresses

obtained high values of σ, while those using small mean stresses obtained much lower values. It is for this reason that Adams and Coker obtained σ for granite not far from 0.25, whereas the Watertown Arsenal obtained values around 0.10.

Summary of Results on Poisson's Ratio.—Each rock studied was stressed under two ranges of loads; hence two values of σ were found. These are listed in table 4, which also gives the respective densities.

As in the case of the measurements of E, the rocks showing the greatest variation of σ with mean stress are seen to be those having the lowest values of σ. Also low values of E accompany low values of σ. This allows one to predict a large compressibility for such rocks at low pressures.

The remarks made on a preceding page on the geophysical significance of the variation in the values of E with mean stress also apply to these results on σ.

The variation of σ with direction in the rock is parallel to that of the change of E.

TABLE 4

YOUNG'S MODULUS, POISSON'S RATIO AND DENSITY OF SOME AMERICAN ROCKS

MATERIAL	SPECIMEN	MEAN STRESS, KG./CM.^{2}	E, DYNES/CM.^{2}	σ	DENSITY, GRAMS/CM.^{3}
Granite	1	3.1	2.71×10^{11}	....	2.61
Quincy surface		78.4	5.08		
Granite		2.8	2.39		
Quincy surface	2	78.4	4.73	....	2.61
Granite		11.2	2.13		
Quincy surface	7	56.0	3.98	....	2.61
Granite		11.2	2.38		
Quincy surface	8	56.0	4.65	....	2.61
Granite		11.2	5.18		
Quincy 100 ft. deep	1	56.0	5.82	....	2.59
Granite		11.2	4.95		
Quincy 100 ft. deep	2	56.0	5.66	....	2.59
Granite		11.2	4.85		
Quincy 100 ft. deep	3	56.0	5.77	....	2.59
Granite		11.2	3.48		
Quincy 235 ft. deep	4	56.0	4.70	....	2.64
Granite		11.2	3.45	0.092	
Quincy 235 ft. deep	5	56.0	4.63	0.130	2.64
Granite	"Hardest Way"	11.2	3.54	0.091	
Rockport 100 ft. deep		56.0	4.26	0.110	2.63
Granite	"Easiest Way"	11.2	3.95	0.082	
Rockport 100 ft. deep		56.0	4.39	0.103	2.63
Norite	1	36.5	8.07	0.212	2.85
Sudbury	2	36.5	7.96	0.224	2.85

TABLE 4 (*continued*)

MATERIAL	SPECIMEN	MEAN STRESS, KG./CM.²	*E*, DYNES/CM.²	σ	DENSITY, GRAMS/CM.³
	3	36.5	8.05	....	2.85
	4	36.5	9.36	....	2.85
Norite		11.2	7.27	0.154	
French Creek	1	56.0	7.67	0.160	3.05
Norite		11.2	5.84	0.108	
French Creek	2	56.0	6.15	0.129	3.05
Norite		11.2	6.72	0.141	
French Creek	3	56.0	7.03	0.146	3.05
Gneiss		0.0	0.328	0.0283	
Pelham (across foliation)	1	9.0	0.43	0.0283	2.64
Gneiss		11.2	1.42		
Pelham (parallel to foliation)	2	56.0	2.20	....	2.64
Gneiss		11.2	1.68	0.082	
Pelham (parallel to foliation)	3	56.0	2.55	0.139	2.64
Diabase		11.2	10.20	0.257	
Vinal Haven	1	56.0	10.15	0.257	2.96
Diabase		11.2	10.25	0.244	
Vinal Haven	2	56.0	10.04	0.244	2.96
Diabase		11.2	10.17	0.261	
Vinal Haven	3	56.0	10.04	0.261	2.96
Marble		11.2	3.43	0.134	
Vermont	1	56.0	4.60	0.180	2.71
Marble		11.2	2.32	0.134	
Vermont (perp. to bedding)	2	56.0	3.99	0.180	2.71
Marble		11.2	3.83	0.142	
Vermont (parallel to bedding)	3	56.0	4.95	0.209	2.71
Limestone		11.2	3.37	0.148	
Pennsylvania	1	56.0	4.57	0.183	2.69
Limestone		11.2	7.60	0.256	
Pennsylvania	2	56.0	7.50	0.286	2.69
Limestone		11.2	6.06	0.168	
Pennsylvania	3	56.0	7.22	0.183	2.69
Quartzitic sandstone		11.2	6.36	0.109	
Pennsylvania	1	56.0	6.65	0.104	2.64
Soft steel		28.0	21.7	0.301	

Description of Rocks Studied in These Experiments.—The following description of the rocks studied is based upon the results of petrographic examinations by H. J. Fraser, E. B. Dane, Jr. (Quincy granite, Rockport granite, French Creek norite, Sudbury norite, Vinal Haven diabase and Pelham gneiss), and C. S. Hurlbut (limestone, sandstone and Vermont marble). The thanks of the writer and of the Harvard Committee on Geophysics are due to these gentlemen.

Granite, Quincy, Massachusetts.—The specimens employed are fairly typical of the main body of the Quincy granite, especially that part of it where field velocity determinations have been made by Leet and Ewing.[3] All the specimens are moderately coarse, holocrystalline and of normal hypiodiomorphic-granular texture. Unusually large thin sections were made from fresh pieces taken at a surface ledge near the J. S. Swingle quarry in West Quincy, and from a depth of 235 feet in the quarry. The volume percentages of the constituent minerals in these two phases are given in the first two columns of table 5.

TABLE 5

VOLUME PERCENTAGES OF THE MINERALS

	QUINCY GRANITE, SURFACE	QUINCY GRANITE, 235 FT. DEEP	ROCKPORT GRANITE 100 FT. DEEP
Quartz	12	26	28
Feldspar:			
Microperthite	65	50	64
Albite	16	13	Small amount
Riebeckite	3	7	3
Soda-amphibole	1	3	3
Magnetite	2	1	..
Titanite, kaolin sericite, leucoxene, limonite, calcite, etc.	1	1	2

In the surface specimen from Quincy, the average diameter of the feldspar and riebeckite grains is about three millimeters; the other specimen has a somewhat smaller grain. The average of Warren's closely similar analyses of the Quincy granite, main phase, can be used to give the chemical composition.[4]

Granite, Rockport, Massachusetts.—In chemical and mineralogical composition this granite is nearly identical with the Quincy granite, while the grain is a little coarser. Table 5, third column, gives the mineral composition of the rock.

Norite, French Creek, Pennsylvania.—This rock, known to quarry men as "St. Peter's trap," is moderately fresh, coarse-grained, holocrystalline and hypidiomorphic-granular. The volume percentages of its constituents are given in table 6.

TABLE 6

NORITE, FRENCH CREEK, VOLUME PERCENTAGES

Plagioclase, $Ab_{35}An_{65}$	51
Augite (75 per cent hedenbergite)	32
Hypersthene	15
Magnetite	0.5
Biotite	0.5
Secondary hornblende, leucoxene, limonite and quartz	1

The diameters of the feldspars reach 1.0 millimeter and average about 0.3 of a millimeter; those of the augite reach 5.5 millimeters and average about 0.8 of a millimeter. The minerals show no sign of strain. No chemical analysis is available.

Norite, Sudbury, Ontario.—On request, the officials of the International Nickel Company, Ltd., kindly presented a large number of diamond-drill cores from the norite at Sudbury, Ontario. Four of these, especially free from alteration, were selected for study. Their original depths below the surface are not known. The respective volume percentages of the constituent minerals are given in table 7.

TABLE 7

NORITE, SUDBURY, VOLUME PERCENTAGES

	CORE 1	CORE 2	CORE 3	CORE 4
Labradorite, $Ab_{48}An_{52}$	..	67	58	70
Andesine, $Ab_{56}An_{44}$	74	..	..	..
Potash feldspar	1	1	2	4
Hypersthene (25 per cent $MgSiO_3$)	15	8	18	10
Augite	3	5	4	2.5
Hornblende	2	5	8	3
Biotite	1.5	6	1.5	3
Quartz	3	8	8	6
Magnetite, apatite, sericite, kaolin, carbonate	0.5	..	0.5	1.5

All four specimens failed to show evidence of any internal strain. Cores 1 and 4 are almost ideally fresh; the other two are slightly altered. The average of five analyses[5] of this norite illustrates fairly well the chemical composition.

Diabase, Vinal Haven, Maine.—This fresh rock from the coast of Maine was chosen as nearly typifying fresh olivine diabase, one of the crystallized products from the quantitatively important basaltic magma. The volume percentages are given in table 8.

TABLE 8

DIABASE, VINAL HAVEN, VOLUME PERCENTAGES

Labradorite, $Ab_{30}An_{60}$	72
Olivine	10
Augite (75 per cent hedenbergite)	10
Hypersthene	2
Biotite	3
Magnetite	2
Uralite, antigorite, apatite, hornblende, quartz, limonite	1

The texture is ophitic. The average lengths of the feldspars, olivines and augites are respectively 0.8 mm., 0.4 mm. and 0.6 mm. The thin

section shows absence of strain. No chemical analysis is available, but the composition must be near that of the average olivine diabase.

Gneiss, Pelham, Massachusetts.—Professor G. W. Bain of Amherst College was good enough to forward a cuboidal block of this fresh, notably foliated, but otherwise homogeneous granitic gneiss, which has an extensive outcrop around Pelham, Massachusetts. Volume percentages of the minerals were measured in sections cut across the foliation and parallel to it. The results are given in table 9.

TABLE 9

GNEISS, PELHAM, MASSACHUSETTS

	PARALLEL TO FOLIATION		ACROSS FOLIATION	
	VOL. PER CENT	AVG. DIAM.	VOL. PER CENT	AVG. DIAM.
Quartz, albite, orthoclase	73	0.2 mm.	84	0.3 mm.
Microcline	13	0.5	8	0.3
Biotite	12	0.7	6	0.25
Apatite	0.5	...	0.5	..
Epidote	1.0	...	1.0	..
Carbonate	0.5	...	0.5	..

As expected, the biotite, quartz and feldspars are elongated more or less systematically in the plane of the foliation. The minerals show sutured boundaries but not much intricate interlocking.

The chemical composition is doubtless well represented by an analysis made of the gneiss at Ward's quarry, Pelham.[6]

Quartzitic Sandstone, Pennsylvania.—Blocks of this Cambrian rock from the Hardyston Quartzite formation, as well as those of the limestone next to be described, were obtained through the courtesy of Professor B. L. Miller of Lehigh University. The sandstone came from the north side of the mountain about midway between Allentown and Bethlehem, Pennsylvania—about 40° 37′ N. Lat., 75° 25′ W. Long. It is a strong, well-cemented rock approaching a typical quartzite in mode of fracture. The rock is comparatively massive, the bedding not being prominent in the specimen studied. The volume percentages, maximum diameters and average diameters of the constituents are given in table 10.

Nearly all of it shows strain shadows in polarized light.

TABLE 10

QUARTZITIC SANDSTONE, PENNSYLVANIA

	VOLUME, PER CENT	MAXIMUM DIAMETER	AVERAGE DIAMETER
Quartz	80	3.0 mm.	0.3 mm.
Microperthite	5	2.0	0.4
Microcline	1	0.6	0.3
Sericite	7	...	..
Magnetite	7	0.5	0.07

Limestone ("*Cement Stone*"), *Pennsylvania.*—This Ordovician rock also represents a formation occurring in Pennsylvania, sampled at the quarry of the Nazareth Cement Company, Nazareth, Pennsylvania—approximately 40° 45.5′ N. Lat., 75° 18′ W. Long. The rock is dark-colored, carbonaceous and fossiliferous. The volume percentages of the minerals are: calcite, 80; carbonaceous material, 17; quartz, 2; feldspar, 1. The grain is extremely fine, the largest grains of calcite having diameters of 0.2 mm. The abundant fossils are bryozoa.

Marble, Vermont.—A large cube of this well-known white marble was kindly presented by the Vermont Marble Company of Proctor, Vermont. The rock is essentially a mass of irregular, interlocking grains of calcite, which almost everywhere in the thin section show pressure-twinning. The only other constituent, quartz, makes up no more than 0.5 per cent of the marble. The carbonate grains average 0.4 mm. in diameter and have a maximum of about 1 mm.

Acknowledgments.—The author wishes to express his sincere thanks to Professor R. A. Daly and Professor P. W. Bridgman for their interest and valuable advice during the progress of the work. It is due to the courtesy of the Harvard University Department of Physics that the facilities of the new research laboratory were made available for these experiments.

* Paper No. 3 published under the auspices of the Committee on Geophysical Research (P. W. Bridgman, R. A. Daly, L. C. Graton, D. H. McLaughlin and H. Shapley) and of the Division of Geological Sciences at Harvard University.

[1] F. D. Adams and E. G. Coker, "Investigation into the Elastic Constants of Rocks," *Publication* **46**, *Carnegie Institution of Washington*, (1906).

[2] C. Bach, "Elastizität und Festigkeit," 4th edition, 64 (1904).

[3] L. D. Leet and W. M. Ewing, "Velocity of Elastic Waves in Granite," *Physics*, **2**, No. 3, March, 160–173 (1932).

[4] C. H. Warren, *Proc. Amer. Acad. Sci.*, **49**, 227 (1913).

[5] A. P. Coleman, E. S. Moore and T. L. Walker, *University of Toronto Studies*, No. **28**, 18 (1929).

[6] B. K. Emerson, *Bull.* **597**, U. S. Geol. Survey, 1917, p. 249.

COMPRESSIBILITY AND ANISOTROPY OF ROCKS AT AND NEAR THE EARTH'S SURFACE*

By W. A. Zisman

Department of Geology and Geography, Harvard University

Communicated May 20, 1933

Purpose of the Investigation.—That remarkably active branch of geophysics, seismology, is dependent for theory and practice on accurate information regarding those physical properties of rocks that are most intimately related to the propagation of elastic waves in the earth. Most important of all for use in diagnosing the composition of hidden rocks is volume compressibility. Due to recent improvements in technique by Bridgman and others, it is now possible to measure the linear and cubic compressibility of a rock under hydrostatic pressures as high as 10,000 to 30,000 kilograms per square centimeter. This is the range of pressure at depths of roughly 35 to 100 kilometers in the earth. While Bridgman has supplied some data on the compressibility of rocks at high pressure, the greater part of the available information comes from the work of Adams and Williamson at the Geophysical Laboratory of the Carnegie Institution of Washington.[1]

Adams and Williamson worked in a pressure range of 2000 to 10,000 "megabars" (now sometimes called "bars"), a megabar being one million dynes per square centimeter and equal to about 1.02 kilograms per square centimeter or about ninety-nine per cent of one atmosphere. Their measurements were made under two different conditions. In the one case the rock specimens were exposed directly to the fluid transmitting the pressure; in the other case the air-dried specimens were covered with thin, impermeable tin-foil which prevented the direct access of the fluid. Adams and Williamson pointed out that the measured compressibility should be the greater in the second case, provided the rock carries cracks or pores in communication. For clearly, when the pressure goes on, the cracks and pores should be more easily closed if the specimen is "covered" than if the transmitting fluid is allowed to enter cracks and pores and so add to the resistance to closure. This theoretical expectation agreed with the results of the many experiments, except in the case of three rocks: marble, Westerly granite and Stone Mountain granite. These anomalies were not explained. During the experiments reported in the present paper, twenty-eight different specimens of rock were tested in both the covered and uncovered states; in every case the behavior was that foretold by Adams and Williamson.

Those authors showed that the compressibility of most rocks decreases rapidly as the pressure rises from an initial amount of three hundred

atmospheres. This fact was discovered when they compared their values of compressibility at pressures over two thousand atmospheres with those obtained by Adams and Coker, who calculated the compressibilities of their rocks from measurements of Young's modulus and Poisson's ratio at a pressure range of about seventy to six hundred atmospheres.[2] From the two sets of results Adams and Williamson constructed curves of compressibility, reaching back to zero pressure, but their method for the lower pressures was by frank extrapolation. For seismological problems relating to rock formations at and close to the earth's surface there has been evident need of a direct experimental study of rocks in the pressure field below that occupied by Adams and Coker. The following paper presents the results of such a study.

Moreover, the geophysicist needs to know how far rocks, especially those at shallow depths, depart from isotropy. That the departure may be great has been shown by recent measurements of Young's modulus and Poisson's ratio at pressures reaching about fifty-six atmospheres.[3] The same fact has become evident in the course of the present research, when cubic compressibilities were determined by first measuring linear compressibility at a pressure range of one to about eight hundred atmospheres. The amount of the anisotropy in standard rocks can be gauged by study of the tables of data to be seen on later pages. It is clear that the seismologist must consider anisotropy when he is dealing with waves passing along the surface layer of the globe.

Description of Piezometer.—The piezometer designed for this work is on the same principle as Bridgman's well-known lever-piezometer used at pressures ranging up to twelve thousand atmospheres.[4] For research in that field of pressure the instrument could manifestly accommodate only short columns of the solids to be investigated. To secure reliable results at low pressures it was necessary to use columns six or eight times as long, the length actually favored being eight inches (twenty centimeters). The width was regularly five-eighths of an inch (1.6 centimeters). Such great lengths and actual thicknesses secured against errors due to the excessive influence of specially large crystals in the rocks, and also assured high sensitivity for the apparatus.

The principle of the Bridgman instrument, built with specifications to match the new kind of investigation, may be briefly reviewed (see Fig. 1). One end of the rock cylinder rests on the piezometer case at the point *A*. The other end is in contact at *B* with the short arm *CB* of a lever *ECB* whose fulcrum is at *C*, the end of the pointed screw *DC*. The arm *EC* is about fourteen times longer than arm *CB*. A spring *JI* maintains *CB* in firm contact with the upper end *B* of the specimen. At *EH* is a stiff straight piece of manganin resistance wire about one centimeter long. This is in gentle, sliding contact with a blade of manganin, *G*, which is

supported but electrically insulated from the main body of the piezometer $M_3M_2M_1$. At F and H are soldered wire connections for the electrical circuit. When the hydrostatic pressure is increased on the liquid in the steel cylinder into which the piezometer is screwed, the specimen shortens in all its dimensions. Point A is fixed, but point B moves downward. Then the lever arm ECB rocks about the fulcrum at C and point E moves to the right. Therefore, the distance FG decreases as the result of the sliding of the manganin wire over the fixed knife-edge contact at G. By measuring the consequent change in electrical resistance of FG, we can, from our knowledge of the resistance per unit length of the wire, calculate the decrease in length of the rock under pressure.

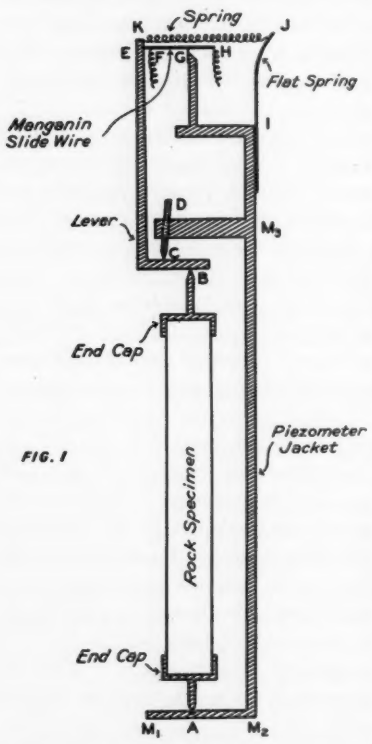


FIG. 1

SCHEMATIC DIAGRAM OF PIEZOMETER

In order to eliminate troubles from the contact resistance at G, the resistance of FG is measured by a potentiometer bridge method described in Bridgman's papers. For this method F and G serve as potential and E and H as current connections.

It is apparent that such an instrument measures only the difference in the hydrostatic linear compressibilities of the rock and the steel of the piezometer case, for when the length of the specimen shortens so does the distance M_2M_3, due to the compressibility of the steel. The true compressibility of the rock is obtained by allowing for that of the steel, for which Bridgman's value, 2.00×10^{-13} cm.^{2}/dyne, was used.

In order to test the reliability of the instrument, a cylinder of soft steel of the same kind as in the piezometer jacket was used in place of the rock specimen. No change in resistance under pressure was observed that could be due to relative motion of the specimen and the jacket, such as would occur if distortion took place in the instrument when the high pressure permitted any internal stresses in the steel to exhibit their presence.

If a specimen of some homogeneous substance free from internal strains and cavities is used, the graph of shortening of the specimen as a function of pressure will be a smooth curve free from hysteresis. This is the most reliable test for the existence of slipping at the bearing points *B* and *C*. A cylinder of pure cast lead was used for such a test. Several independent runs gave the same result and proved the absence of hysteresis. At thirty degrees centigrade the linear compressibility of lead was found to be 8.40×10^{-13} cm.^{2}/dyne. Bridgman[5] obtained 8.07×10^{-13} cm.^{2}/dyne. This difference may be due to the lack of perfect annealing in my specimen.

The hydrostatic pressure was measured with a manganin resistance coil by inserting a 250-ohm coil of well seasoned manganin wire in the pressure chamber and measuring its change of resistance under pressure with a Carey-Foster bridge. The pressure calibration of this type of secondary gauge is simple. The resistance of manganin was shown by Bridgman to change linearly with pressure in the range 0–10,000 kg./cm.^{2}. Therefore it is necessary only to measure the resistance of the coil at atmospheric pressure and again at some higher known pressure, such as that at which mercury freezes at zero centigrade (7640 kg./cm.^{2}). There was no change in the calibration of the coil used during the year needed for these measurements.

Preparation of Specimens.—The rock cylinders were core-drilled from the same blocks as those from which were cut the two-inch-diameter cylinders used for the measurement of Young's modulus and Poisson's ratio described in a previous paper.[3] The same method of indicating the directions of coring of the cylinder for compressibility tests was used. The cores were ground carefully to form accurate right cylinders.

The compressibility of the specimen was measured when immersed in kerosene, the fluid transmitting the pressure. This will be referred to as the compressibility of the "uncovered" rock. In contrast is the compressibility of the "covered" rock, the cylinder in this case having been enclosed in an impermeable sheath of 0.002-inch annealed copper foil. This covering was strong enough to prevent entrance of the kerosene into the rock and flexible enough to transmit the pressure freely to the rock within.

The method of preparing these covers may be indicated. The rock cylinder was rolled in a sheet of foil, much as the tobacco of a cigarette is rolled in its paper cover. The end of the rolled sheet of foil then formed a seam running the length of the cylinder. This seam was soft-soldered, and to facilitate the operation the foil edge was tinned before rolling. Thin caps were made for covering the ends of the cylinder, by spinning a sheet of copper into a cap whose inside diameter was five-eighths of an inch and whose length was from a quarter to three-eighths of an inch. The cap fitted snugly over the end of the specimen and the rim was soft-soldered to the copper covering beneath. When hydrostatic pressure was

applied, the excess foil present was forced into the surface cavities in the rock, making an extremely close fit like that of the skin of an onion.

After being covered, the rock was exposed to a pressure of 1000 kg./cm.^{2} for five minutes. Then a small hole was made in the cover. If there had been a leak, it was detected by the issuing of small bubbles of kerosene from this hole. After some practice in the technique it was found that few covers leaked. After this test, the small hole was sealed with soft solder, and the specimen was ready for a compressibility measurement. Upon

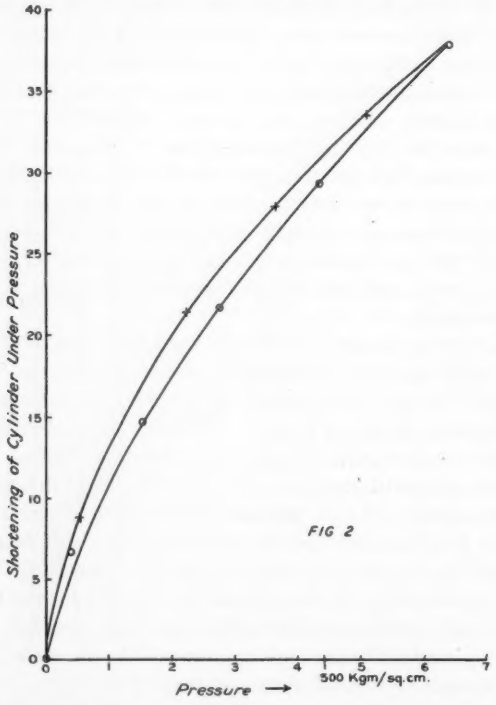


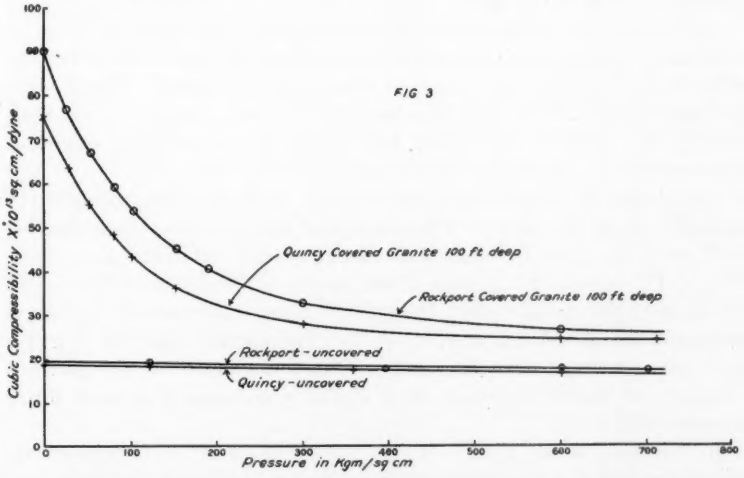
FIG 2

completion of an experiment the foil was stripped from the rock to permit detection of further evidence of a leak.

A blank run on the compressibility of a cylinder of soft steel covered with foil was made, and it was thus possible to prove that there was no contribution to the observed compressibility attributable to the existence of air spaces between the foil and the flat ends of the rock. When the compressibility of the covered rock was measured, the smallness of the permanent set found after each pressure run was considered good reason for believing the observed pressure effect to be due to the change in length

of the rock rather than to the squeezing of the foil covering against the ends of the specimen. This conclusion was strengthened by the fact that successive runs on the same rock using entirely new covers gave the same compressibility within a few per cent. Moreover, the compressibility of covered and uncovered obsidian glass was found to agree, as it should, since such material is non-porous.

Description of Results.—Figure 2 is the graph of the shortening of a cylinder of covered Rockport granite against pressure. There is a permanent set amounting to about three per cent of the total shortening. The uncovered rocks show somewhat smaller amounts of permanent set. From such a graph it is not difficult to obtain measurements of the compressibility as a function of pressure by reading off the slope for, say, eight values of the pressure.



Just before taking a compressibility measurement to 800 kg./cm.^{2} a ten-minute pressure seasoning to 1000 atmospheres was given the apparatus and specimen. A repetition of the compressibility measurement was found unnecessary. Even when the pressure seasoning lasted as long as five hours, the compressibility was unchanged.

The permanent set was found greatest for the rocks exhibiting the largest change in compressibility with pressure, that is, for the most porous rocks. On the other hand, pronounced hysteresis, as indicated by a great width for the hysteresis loop, was not always easily linked with high porosity. It is quite possible that most of the hysteresis was due to the presence of internal strains in the rocks, such as there must always be in a rock essentially composed of non-cubic crystals.

In the following tables the linear (α) and cubic compressibilities (β) at thirty degrees centigrade are given as a function of pressure for both the covered and uncovered rocks. The cubic compressibility was computed by tripling the mean value of the linear compressibilities. In figure 3 is presented the cubic compressibility of granite as a function of pressure.

The difference in behavior of each rock with and without cover is striking. The values for the uncovered rocks should be considered as the more accurately determined, chiefly because the small change in compressibility with pressure then observed made it easier to obtain from the graphs accurate values of shortening with pressure. The values of the compressibilities for the compact rocks are good to two per cent, while for the rocks showing rapid changes in compressibility with pressure, the error may rise as high as five per cent for the initial compressibility and two to three per cent at 840 kg./cm.^{2}.

Throughout the whole series of measurements the compressibility of each covered rock is greater than its value when uncovered. The difference between the two becomes smaller as the pressure is raised. This is what one would expect, for at high pressures the looseness of texture in the rock due to the presence of cavities and cracks is substantially removed. Thus we find the two values almost equal at 840 kg./cm.^{2}.

The change in the linear compressibility with direction parallels the vectorial change of E and σ. This is a natural relation, the cylinders having been cut respectively from the same blocks of rock and in the same directions. The results for French Creek norite will serve as an example. Of the three directions studied, the greatest linear compressibility occurs in specimen No. 2, and the least in No. 1. The measurements of E and σ show number two to have the smallest value and number one the greatest. Thus, as we should expect, a large E and σ accompany a small linear compressibility.

Description of Rocks Used.—Most of the rocks used in these experiments were described in a previous publication.[3] The following descriptions of additional rocks studied here are due to the kindness of H. J. Fraser, E. B. Dane, Jr., and C. S. Hurlbut, of the Department of Geology of Harvard University.

Obsidian, Lipari Islands.—This well-known material is essentially an unstrained rhyolitic glass, containing thread-like crystallites of iron oxide and minute feldspars totaling about two per cent of the volume. Alignment of the crystallites indicates a definite flow structure, which is macroscopically not conspicuous. A chemical analysis of this glass is to be found in Washington's Tables.

Peridotite (Dunite), North Carolina.—A large block of this astonishingly fresh, ultrabasic rock was generously supplied by Mr. A. W. Hyatt of Balsam (Balsam Gap), North Carolina. A sugary, holocrystalline, allo-

TABLE 1

LINEAR (α) AND CUBIC COMPRESSIBILITIES (β) AT 30°C. (CM.²/DYNE)
(Cylinders numbered, mutually at right angles to one another)

	PRESSURE (KG./CM.²)	$\alpha \times 10^{13}$ CYL. 1	$\alpha \times 10^{13}$ CYL. 2	$\alpha \times 10^{13}$ CYL. 3	$\beta \times 10^{13}$ MEAN
Quincy granite at surface (uncovered)	0	6.92	9.42	8.32	24.7
	60	6.76	9.05	8.12	24.0
	120	6.62	8.69	7.93	23.2
	240	6.30	8.27	7.62	22.2
	360	6.10	7.95	7.40	21.4
	480	5.93	7.78	7.25	21.0
	600	5.89	7.74	7.15	20.8
	720	...	7.73	7.09	20.7
	840	...	7.72	7.07	...
Quincy granite, 100 feet deep (covered)	0	20.9	31.1	23.6	75.6
	24	18.9	25.3	20.4	64.5
	48	17.1	21.1	17.4	55.6
	72	15.8	17.4	15.5	48.7
	96	14.7	15.1	14.2	43.9
	120	13.6	13.7	13.0	40.2
	144	12.7	12.7	12.1	37.6
	180	11.5	11.6	11.2	34.4
	300	9.52	9.81	9.51	28.8
	600	8.26	8.54	8.44	25.3
	720	8.18	8.22	...	24.6
Quincy granite, 100 feet deep (uncovered)	0		6.50	6.23	19.2
	60		6.35	6.15	18.8
	120		6.20	6.07	18.5
	240		5.88	5.93	17.7
	360		5.72	5.82	17.3
	480		5.58	5.72	17.0
	600		5.51	5.65	16.7
	720		5.49	5.63	16.7
	840		5.48	5.61	16.6
Quincy granite, 235 feet deep (covered)	0	30.0			89.9
	24	26.3			78.8
	48	23.2			69.7
	72	20.8			62.3
	96	18.7			56.0
	120	16.9			50.8
	144	15.5			46.5
	180	13.8			41.3
	300	10.4			31.2
	600	8.32			25.0
	720	7.84			23.6
Quincy granite, 235 feet deep (uncovered)	0	7.68	6.35		21.1

TABLE 1 (*Continued*)

	PRESSURE (KG./CM.²)	CYL. 1 $\alpha \times 10^{13}$	CYL. 2 $\alpha \times 10^{13}$	CYL. 3 $\alpha \times 10^{13}$	$\beta \times 10^{13}$ MEAN
	24	7.44	...		...
	60	7.09	6.26		20.1
	120	6.71	...		...
	180	6.38	...		...
	240	6.22	5.91		18.2
	360	6.15	5.71		17.8
	480	...	5.49		...
	600	...	5.31		...
	720	...	5.13		...
Rockport granite, 100 feet deep (covered)	0	33.0	28.2		91.7
	24	26.7	25.5		78,2
	48	22.1	23.4		68.2
	72	18.8	21.3		60.2
	96	16.8	19.6		54.6
	120	15.3	18.3		50.4
	144	14.2	17.1		47.0
	180	12.7	15.4		42.2
	300	10.6	11.5		33.2
	600	8.45	9.29		26.6
Rockport granite, 100 feet deep (uncovered)	0	6.46	6.53		19.5
	60	6.33	6.35		19.1
	120	6.22	6.22		18.7
	240	5.99	6.02		18.0
	360	5.92	5.91		17.7
	480	5.86	5.87		17.5
	720	5.79	5.75		17.3
French Creek norite (covered)	0	17.7	22.1	19.1	59.0
	24	16.1	21.1	17.4	54.6
	48	14.7	20.2	15.9	50.8
	72	13.5	19.1	14.6	47.1
	96	12.5	17.9	13.5	43.9
	120	11.5	17.0	12.5	41.1
	144	10.8	15.9	11.7	38.4
	180	9.74	14.4	10.6	34.8
	300	7.24	10.9	7.82	25.9
	600	4.74	6.42	5.40	16.6
	720	4.48	5.78	4.93	15.2
French Creek norite (uncovered)	0	4.58	4.94	4.43	14.0
	60	4.54	4.60	4.42	13.6
	120	4.52	4.42	4.38	13.4
	240	4.45	4.30	4.33	12.9
	360	4.39	4.22	4.30	12.9
	480	4.33	4.19	4.26	12.7
	600	4.26	4.17	4.24	12.6
	720	4.18	4.16	4.22	12.5
	840	4.12	4.15	4.20	12.5

TABLE 1 (*Continued*)

	PRESSURE (KG./CM.²)	CYL. 1 $\alpha \times 10^{13}$	CYL. 2 $\alpha \times 10^{13}$	CYL. 3 $\alpha \times 10^{13}$	$\beta \times 10^{13}$ MEAN
Sudbury norite (covered)	0	10.2	10.8		31.5
	24	9.58	10.16		29.6
	48	8.97	9.41		27.6
	72	8.38	8.80		25.8
	96	7.77	8.22		24.0
	120	7.17	7.75		22.4
	144	6.69	7.40		21.1
	180	6.07	6.93		19.5
	300	5.50	5.96		17.2
	600	5.40	5.57		16.5
	720	5.34	5.56		16.3
Sudbury norite (uncovered)	0	5.60	5.42*		16.5
	60	5.51	5.42*		16.4
	120	5.44	5.42*		16.3
	240	5.33	5.42*		16.1
	360	5.26	5.42*		16.0
	480	5.20	5.42*		15.9
	600	5.18	5.42*		15.7
Olivine diabase (covered)	0				17.1
	120				15.9
	300				14.3
	480				13.3
	720				12.6
Olivine diabase (uncovered)	0–200				14.6
	400–720				13.3
Peridotite (uncovered)	0	4.50	2.98*		11.2
	60	4.32	2.98*		10.9
	120	4.20	2.98*		10.9
	240	3.89	2.98*		10.3
	360	3.68	2.98*		9.98
	480	3.48	2.98*		9.68
	600	3.36	2.98*		9.51
	720	3.28	2.98*		9.37
Orthogneiss (covered; cylinder No. 1 cut with axis at right angle to plane of schistosity)	0			48.5	
	120			27.4	
	300			14.3	
	600			9.61	
	720			9.06	
Orthogneiss (uncovered)	0	8.24	6.32	6.18*	20.7
	60	7.62	6.17	6.18*	20.0
	120	7.13	6.03	6.18*	19.3
	240	6.36	5.82	6.18*	18.4
	360	5.85	5.65	6.18*	17.7
	480	5.45	5.57	6.18*	17.2
	600	5.10	5.52	6.18*	16.8
	720	...	5.48	6.18*	...

* No sensible change of α in range of pressure.

TABLE 1 (*Continued*)

	PRESSURE (KG./CM.²)	CYL. 1 $\alpha \times 10^{13}$	CYL. 2 $\alpha \times 10^{13}$	CYL. 3 $\alpha \times 10^{13}$	$\beta \times 10^{13}$ MEAN
Quartzitic sandstone (covered; cylinder No. 1 cut with axis at right angle to bedding)	0	19.7	23.3	15.7	58.7
	24	18.3	21.5	14.3	54.1
	48	17.2	19.7	13.3	50.2
	72	16.3	18.3	12.4	47.1
	96	15.6	17.3	11.8	44.8
	120	15.0	16.5	11.3	42.8
	144	14.5	15.9	10.9	41.3
	180	13.8	15.1	10.4	39.2
	300	12.1	13.2	9.48	34.8
	600	10.9	11.0	8.92	30.9
	720	10.5	10.5	8.82	29.9
Quartzitic sandstone (uncovered)	0	8.39*	9.40	8.86*	26.7
	60	8.39*	9.31	8.86*	26.6
	120	8.39*	9.22	8.86*	26.5
	240	8.39*	9.07	8.86*	26.3
	360	8.39*	8.92	8.86*	26.2
	480	8.39*	8.82	8.86*	26.1
	600	8.39*	8.76	8.86*	26.0
	720	8.39*	8.74	8.86*	26.0
Vermont marble (covered cylinder No. 2 cut with axis at right angle to bedding)	0	58.8	75.0	46.5	180.0
	24	42.0	47.1	26.2	115.0
	48	24.0	26.1	18.0	68.2
	72	18.0	18.6	13.6	50.3
	96	14.2	15.9	10.9	40.9
	120	11.7	12.2	9.12	33.1
	144	10.2	10.6	8.32	29.2
	180	8.79	9.12	7.95	25.9
	300	6.67	7.25	6.59	20.5
	600	4.29	5.62	5.10	15.0
	720	...	5.46	...	...
Vermont marble (uncovered)	0	4.17	5.19*	4.53	13.9
	120	4.06	5.19*	4.47	13.8
	240	3.90	5.19*	4.40	13.5
	360	3.79	5.19*	4.28	13.3
	480	3.69	5.19*	4.09	12.9
	600	3.61	5.19*	3.82	12.6
	720	3.53	5.19*	3.50	12.2
	840	3.46	5.19*	...	...
Limestone (covered; cylinder No. 1 cut with axis at right angle to bedding)	0	14.5	6.20	8.49	29.2
	60	13.9	6.12	8.29	28.2
	120	13.4	6.02	8.09	27.5
	240	12.4	5.90	7.75	26.1

* No sensible change of α in the range of pressure.

TABLE 1 (*Continued*)

	PRESSURE (KG./CM.²)	CYL. 1 $\alpha \times 10^{13}$	CYL. 2 $\alpha \times 10^{13}$	CYL. 3 $\alpha \times 10^{13}$	$\beta \times 10^{13}$ MEAN
	360	11.6	5.79	7.44	24.9
	480	11.0	5.73	7.25	23.9
	600	10.7	5.68	7.10	23.5
	720	10.4	5.65	6.94	23.0
Limestone (uncovered)	0	11.6*	5.88*	7.22	24.7
	60	11.6*	5.88*	7.13	24.6
	120	11.6*	5.88*	7.06	24.5
	240	11.6*	5.88*	6.96	24.5
	360	11.6*	5.88*	6.81	24.2
	480	11.6*	5.88*	6.71	24.1
	600	11.6*	5.88*	6.58	24.1
	720	11.6*	5.88*	6.45	24.0
	840	11.6*	5.88*	6.32	23.7
Dolomite (massive; covered)	0	14.2	10.60		37.1
	24	13.0	10.2		34.8
	60	10.6	9.63		30.4
	120	8.28	8.65		25.4
	240	6.62	6.73		20.1
	360	5.88	5.37		16.9
	480	5.40	4.91		15.5
	600	5.18	4.70		14.8
	720	4.99	4.44		14.2
Dolomite (uncovered)	0	4.11	3.81		11.9
	360	4.11	3.81		11.9
	720	4.11	3.81		11.9

* No sensible change of α in range of pressure.

triomorphic mass of medium grain, it is composed of olivine (92 per cent Mg_2SiO_4 and 8 per cent Fe_2SiO_4, as measured optically), 97 per cent; serpentine, two per cent; with chromite, carbonate and hornblende totaling one per cent. The olivines reach 4.4 mm. in diameter, the average being 0.75 mm. The chemical composition of this dunite must approximate that of the olivine alone, as found by Gonyer.[6]

Dolomite, Pennsylvania.—This material of Cambrian age, obtained through the courtesy of Professor B. L. Miller of Lehigh University, came from the Bethlehem Mines Corporation quarry just east of Bethlehem, Pennsylvania (about 40° 37′ N. Lat., 75° 22′ W. Long., Allentown topographic sheet). The rock is massive, with obscure bedding in the block actually studied. The mineral data are as follows:

	VOLUME, PER CENT	MAXIMUM DIAMETER, MM.	AVERAGE DIAMETER, MM.
Calcite	98	0.25	0.01
Quartz	1	0.25	0.1
Feldspar	1	0.2	0.1

TABLE 2

SUMMARY OF COMPRESSIBILITY DATA

	PRESSURE (KG./CM.²)	COMPRESSIBILITY $\beta \times 10^{13}$ UNCOVERED	COVERED	DENSITY
Quincy granite, at surface	0	24.7	..	2.61
	720	20.7	..	
Quincy granite, from depth of 100 feet	0	19.2	76.6	2.59
	720	16.7	24.6	
Quincy granite, from depth of 235 feet	0	21.1	89.8	2.64
	720	17.8[1]	23.6	
Rockport granite	0	19.5	91.7	2.63
	720	17.3	26.6[1]	
French Creek norite	0	14.0	59.0	3.05
	720	12.5	15.2	
Sudbury norite	0	16.5	31.5	2.85
	720	15.7[1]	16.3	
Olivine diabase	0	14.6[2]	17.1	2.96
	720	13.3[3]	12.6	
Peridotite	0	11.2	..	3.27
	720	9.37	..	
Orthogneiss	0	20.7	..	2.64
	720	16.6	..	
Obsidian	0	29.6	29.5	2.34
	720	29.6	29.5	
Quartzitic sandstone	0	26.7	58.7	2.64
	720	26.0	29.9	
Marble	0	13.9	180.0	2.71
	720	12.2[1]	15.0	
Limestone	0	24.7	29.2	2.69
	720	24.0	23.0	
Dolomite	0	11.9	37.1	2.82
	720	11.9	14.2	
Soft steel	0	5.99	..	7.79

[1] Value at 600 kg./cm.².

[2] Mean for range of 0–200 kg./cm.².

[3] Mean for range of 400–720 kg./cm.².

Numerous cracks carry a limonitic stain and along them the calcite has been locally recrystallized, with an increase in the size of the crystals.

Summary.—The compressibilities of a number of rocks of geophysical interest have been measured at thirty degrees centigrade for pressures up to 840 kg./cm.². The prediction of Adams and Williamson of a rapid rise in compressibility as the pressure is decreased has been verified. It is a remarkable effect in the case of covered rocks, and a very much smaller one for uncovered rocks. The variation of the compressibility with pressure is greater for the more porous rocks. Thus, in order of decreasing magnitude of this effect, we have marble, granite, limestone, Sudbury norite and Vinal Haven diabase. The compressibility of the uncovered rock is always found to be less than that of the same rock when

covered. At pressures up to 300 kg./cm.^{2} the difference may be large, but it rapidly approaches zero at higher pressures. In many cases the difference is already unimportant at 840 kg./cm.^{2}. Simple explanations of these pressure effects can be given, based upon the existence of many easily compressed cracks and cavities between mineral grains. The linear compressibility often shows a large variation with direction in the rock; it is naturally most striking in layered or bedded rocks.

Acknowledgments.—These experiments, suggested by the Committee on Geophysics of Harvard University, were performed in the Research Laboratory of Physics, through the courtesy of the Department of Physics of the same institution. I wish to thank also Professor P. W. Bridgman and Professor R. A. Daly for their active interest and advice.

* Paper No. 4 published under the auspices of the Committee on Geophysical Research and the Division of Geological Sciences at Harvard University.

[1] P. W. Bridgman, *Amer. Jour. Sci.*, **7**, 81 (1924); L. H. Adams and E. D. Williamson, *Jour. Franklin Inst.*, **195**, 475 (1923); *Smithsonian Report for 1923*, 241 (1925); *Proc. Nat. Acad. Sci.*, **12**, 275 (1926); *Gerlands Beitr. z. Geophys.*, **31**, 315 (1931); *Jour. Wash. Acad. Sci.*, **21**, 381 (1931).

[2] F. D. Adams and E. G. Coker, "Investigation into the Elastic Constants of Rocks," *Publication* **46**, *Carnegie Institution of Washington* (1906).

[3] W. A. Zisman, "Young's Modulus and Poisson's Ratio with Reference to Geophysical Applications," these PROCEEDINGS, **19**, 653–665 (1933).

[4] P. W. Bridgman, *Proc. Amer. Acad. Sci.*, **56**, 3, February, 64–67 (1921).

[5] P. W. Bridgman, *Ibid.*, **58**, 5, Jan., 197 (1923).

[6] Ross, Shannon and Gonyer, *Econ. Geol.*, **23**, 545 (1928).

COMPARISON OF THE STATICALLY AND SEISMOLOGICALLY DETERMINED ELASTIC CONSTANTS OF ROCKS*

BY W. A. ZISMAN

DEPARTMENT OF GEOLOGY AND GEOGRAPHY, HARVARD UNIVERSITY

Communicated May 20, 1933

The speeds, V and v, of longitudinal and transverse waves of sound in an infinite, homogeneous and isotropic medium obeying Hooke's law are related to the other elastic constants in a simple manner. If Young's modulus, the cubic compressibility, Poisson's ratio and the density are E, β, σ and ρ, respectively, then it has been shown by the mathematical theory of elasticity that these six constants are related as follows:

$$V = \sqrt{\frac{E(1-\sigma)}{\rho(1+\sigma)(1-2\sigma)}} \tag{1}$$

$$\frac{V}{v} = \sqrt{\frac{2(1-\sigma)}{1-2\sigma}} \tag{2}$$

$$\beta = \frac{3(1-2\sigma)}{E}. \tag{3}$$

In short, we can regard such a medium as having its elastic behavior completely described when we assign the density and any two of the constants E, σ, β, V and v.

It is fundamental for the theory of the seismological analysis of the structure and composition of the earth's interior that we investigate experimentally the validity of assuming that rocks do behave elastically as homogeneous elastic media obeying Hooke's law.

There are a number of methods of studying the general elastic behavior of rocks. We can measure the elastic constants E and σ with a static testing machine. This will disclose any deviations from Hooke's law. If rock specimens of different orientations are cut from a massive block of freshly quarried rock, we can ascertain the amount of elastic anisotropy. With the static constants so obtained the compressibility can be calculated from equation (3). Comparison of this computed value with the compressibility measured directly for the same solid by subjecting it to hydrostatic pressure furnishes a sensitive test of any departure from homogeneity and isotropy. Again, we can measure V and v seismologically in the quarry where the rock specimens for the static tests were obtained, and from equations (1) to (3) compute E, σ and β. These may be appropriately called the dynamical elastic constants. A comparison of these with the static values is a manifest necessity, if the seismologist is to use the elasticity as a criterion of the nature of rocks below the earth's surface.

With the support of the Harvard Committee on Geophysics, the author commenced in 1931 a laboratory investigation of the elastic properties of rocks, including some in which the seismic-wave velocities at and near the surface have been measured.

The purpose of this paper is to interpret the related results of all these experimental studies.

The laboratory measurements have already been published.[1,2] Leet and Ewing have published dynamical elastic constants for Quincy granite[3] and a report on those for Rockport granite and Sudbury norite is in preparation.

The speeds observed at Quincy were $V = 4.96$ km./sec. and $v = 2.48$ km./sec. From these, the dynamical constants were found to be

$$E = 4.3 \times 10^{11} \text{ dynes/cm.}^2$$
$$\sigma = 0.33$$
$$\beta = 22.7 \times 10^{-13} \text{ cm.}^2\text{/dyne}$$

with probable errors of about 2 per cent.

The statical measurements of E and σ and the linear compressibility measurements on the granites from Quincy, Westerly and Rockport all demonstrated their anisotropy. This became particularly evident in the case of the Quincy granite of which many samples were studied. In all instances the values of E for even almost perfectly fresh surface rock were found to be smaller than those for fresh rock taken from depths of 100 and 235 feet. The measurements of σ show little difference with samples or locality. The mean values for E and σ in Quincy were 3.5×10^{11} dynes/cm.^{2} and 0.10, respectively. Thus the statical values are definitely smaller than the dynamical.

In the author's report on the measurement of the hydrostatic compressibilities of rocks, two values of the compressibility were distinguished: the compressibility of the covered rock and that of the uncovered rock. The former was found by studying the change in volume of the rock when covered by an impermeable cover of metal foil. The latter was found by studying the changes in volume of the rock when exposed directly to the hydrostatic fluid so that the liquid was able to penetrate into the pores of the rock. The former value always exceeds the latter. The difference between the two values of compressibility is least for the compact rocks and greatest for the most porous. As the hydrostatic pressure increases, the two values of the compressibility approach each other, the difference usually becoming small at pressures around 800 kg./cm.^{2}.

This contrasted behavior is due to the existence of cracks and cavities among the crystals making up the rocks.

The mean value of the initial compressibility of covered Quincy granite is 90×10^{-13} cm.^{2}/dyne. The mean initial value for the uncovered

granite is 21.6×10^{-13} cm.^{2}/dyne. Evidently, the compressibility of the uncovered rock agrees well with the seismologically determined value.

Summarizing the results for granite, we find the statical constants E and σ are considerably smaller than the dynamical; the statical compressibility of the covered rock is much greater than the dynamical value; and the statical compressibility of the uncovered rock is in good agreement with the dynamical value.

In the norite at Sudbury, Ontario, Leet obtained accurate measurements on elastic wave speeds. The results were $V = 6.22$ km./sec. and $v = 3.49$ km./sec. The density being 2.85, these give:

$$E = 8.82 \times 10^{11} \text{ dynes/cm.}^2$$
$$\sigma = 0.27$$
$$\beta = 15.6 \times 10^{-13} \text{ cm.}^2\text{/dyne}$$

The mean statical values are:

$$E = 8.36 \times 10^{11} \text{ dynes/cm.}^2$$
$$\sigma = 0.22$$

The initial compressibility of the covered rock is 28.6×10^{-13} cm.^{2}/dyne, and of the uncovered rock 16.4×10^{-13} cm.^{2}/dyne. Thus the statically determined constants E and σ in norite, though in better agreement with the dynamical values than was the case in granite, are again definitely smaller than the dynamical. The compressibility of uncovered rock is in good agreement with the dynamically determined value.

The Sudbury norite is more compact than Quincy granite; this is indicated by the much higher value of Young's modulus for norite.

These two comparisons of field and laboratory work suggest the following generalizations: The statically determined elastic constants E and σ are smaller than the respective seismological or dynamical values; the compressibility of uncovered rock is essentially the same as the seismically effective compressibility; the more compact the rock the more nearly do the statical and dynamical constants agree, and the nearer together are the respective compressibilities of covered and uncovered specimens.

Recently Richards published the results of his measurements on the statical and seismological elastic constants of limestone cores cut from various depths in a large oil-bearing structure in southwest Persia.[4] As he was mainly concerned with seismic waves traveling in directions parallel to the bedding of the rock, he confined his statical measurements to those giving the elastic constants E and σ in the plane of the bedding. The mean statical values of E and σ were 5.7×10^{11} dynes/cm.^{2} and 0.25, while the mean seismological values were 5.5×10^{11} dynes/cm.^{2} and 0.29, respectively. The average compressibility computed from his seismological measurements is 22.8×10^{-13} cm.^{2}/dyne.

The Persian limestone is closely similar to a compact Pennsylvania limestone studied at Harvard, for which, when covered, the observed compressibility is 29.0×10^{-13} cm.^{2}/dyne and, when uncovered, 24.7×10^{-13} cm.^{2}/dyne. The small observed difference between the two types of compressibility clearly indicates great compactness of the rock. The measurements on the linear compressibility for directions perpendicular and parallel to the plane of bedding of the Pennsylvania limestone proved strong anisotropy. This was shown also by measurements of E and σ on specimens cut in and at right angles to the plane of the bedding. The mean values of E and σ in the plane of bedding were 6.7×10^{11} dynes/cm.^{2}. and 0.25, respectively. Thus my statical results are in fair agreement with Richards' statical values.

As in the comparisons with granite and norite, we find bedded limestone exhibiting a lower statical value of σ than that obtained seismologically, and here too the compressibility of the rock, when uncovered, is nearly that deduced from the seismic waves.

The departure of the rocks from ideal elasticity may be tested in another way. Table 1 gives the results of computing from the values of E and σ and from equation (3) the compressibilities at two different pressures, on the arbitrary assumption that all the rocks behave as homogeneous, isotropic media. In nearly all cases the computed compressibility is smaller than the observed compressibility, the difference increasing with the porosity of the material. The marble is the most porous; the Vinal Haven diabase the least.

TABLE 1

ROCK	COMPRESSIBILITY AT 11.2 KG./CM.^{2}		COMPRESSIBILITY AT 56.0 KG./CM.^{2}	
	CALCULATED CM.^{2}/DYNE	OBSERVED CM.^{2}/DYNE	CALCULATED CM.^{2}/DYNE	OBSERVED CM.^{2}/DYNE
Limestone, Pennsylvania	32.6×10^{-13}	29.0×10^{-13}	36.4×10^{-13}	28.3×10^{-13}
Diabase, Vinal Haven	14.5	17.0	14.3	16.5
Norite, French Creek	33.2	58.0	30.7	48.0
Granite, Rockport 100 ft.	66.2	84.0	54.5	65.0
Granite, Quincy 235 feet	70.6	86.0	47.5	67.0
Marble, Vermont	71.0	140.0	41.4	59.0
Quartzitic sandstone, Pennsylvania	36.9	56.4	35.7	48.7

ROCK	COMPRESSIBILITY AT 36.5 KG./CM.^{2}	
	CALCULATED CM.^{2}/DYNE	OBSERVED CM.^{2}/DYNE
Norite, Sudbury	20.2×10^{-13}	28.6×10^{-13}

However, there is no theoretical justification for assuming anything like equivalence for the compressibilities, computed and observed, at such a pressure as 11.2 kg./cm.^{2}. Possibly if the computed value for this pressure be compared with the observed hydrostatic compressibility at three times as much, or 33.6 kg./cm.^{2}, a more significant result might be

obtained. Even then the respective differences are considerable and greater than can be explained by experimental error. It is further noteworthy that the differences are not due to lack of isotropy in the rocks, Vinal Haven diabase and Quincy granite at 235 feet, for example, being almost ideally isotropic.

The analogous values of Poisson's ratio may also be compared. Again assuming the validity of equation (3), and also assuming the correctness of the observed initial values for compressibility and Young's modulus, we compute σ for the rocks in both the covered and uncovered conditions, and at zero pressure. The results are shown in table 2. The values given in the right hand column are tending toward those expected if the rocks were free from cracks and pores. Although these were by no means all closed by the actual pressures, the computed values of σ are not wildly unreasonable. In utter contrast are the values given in the first column of table 2. Manifestly the classic equations cannot be applied to most rocks of the earth's crust except when these are at such depths as to ensure the closing of cracks and pores.

TABLE 2

ROCK	POISSON'S RATIO USING β, COVERED	POISSON'S RATIO USING β, UNCOVERED
Granite, Quincy 100 ft.	−0.11	+0.31
Granite, Rockport 100 ft.	−0.06	+0.38
Granite, Quincy 235 ft.	+0.03	+0.39
Diabase, Vinal Haven	+0.21	+0.25
Norite, French Creek	−0.14	+0.35
Norite, Sudbury	+0.06	+0.27
Quartzitic sandstone, Pennsylvania	−0.12	+0.22
Limestone, Pennsylvania	+0.24	+0.28
Marble, Vermont	−0.36	+0.43

The probable cause of the difference between the statical and dynamical compressibilities of standard rocks at low pressure may now be considered.

Let us assume that the dynamical methods, of which the seismological method is but one, give values of E, σ and β for the rock when free from cracks and easily compressed cavities. The statical values of E and σ are lower because part of the yielding of the rock is increased by deformation of cracks and cavities. The compressibility of the covered rock is high for the same reason. The compressibility of the uncovered rock is the same as the dynamically obtained compressibility, because the uncovered rock is soaked with the fluid transmitting the hydrostatic pressure and hence the measured effect is only that due to the compressibility of the solid material in the rock.

It is desirable, if possible, to give a physical foundation for the generalization made above, which, after all, is an inference based on a limited number of comparisons of seismological and statical measurements.

A qualitative explanation can be based upon considerations of wave energy. A plane wave of sound is propagated in the direction indicated by the arrows in figure 1. The wave is supposed to pass over two cracks A and B, such as might mark the boundary between adjacent crystals in this rock. The portion of the wave striking A and B would suffer loss of energy, due to the two reflections at the rock-air and air-rock boundaries. A small amount of wave energy would be transmitted across the boundary and a large amount would be scattered around the cavities A and B. As the cavities in the rock will not in general form an orderly array, we realize that such scattered energy will be of small importance except as a cause of absorption for the wave propagation over long distances in the rocks. Thus we are led to conclude that most of the wave energy transmitted through the rock masses will be made up of the beams transmitted between the crevices and cavities. But these beams will move with the velocity corresponding to the compact rock material. In other words the seismically effective elastic constants will be the high E and σ one would obtain in a statical measurement on a rock of that mineralogical composition, free from crevices and cavities. For waves of low frequency the elastic constants will be the statical values, and for high frequencies they will be the high values due to a rock free from cracks and cavities.

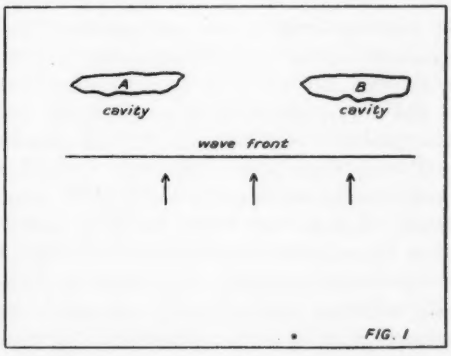


FIG. 1

Recently, Dr. D. S. Muzzey, working in the Research Laboratory of Physics at Harvard University, has obtained dynamical measurements of E for various rocks by using a magnetostriction method to set the specimens vibrating. The same five-eighths-inch rock cylinders were used that were employed for my measurements of the hydrostatic compressibilities. Thanks to the kindness of Dr. Muzzey, it is possible to present here some of his results. A detailed description of method and results will be published separately by him. The following data in table 3 are for sound frequencies ranging from 9000 to 13,000 cycles per second.

The experimental error is estimated to be plus or minus two per cent. Dr. Muzzey encountered much difficulty with the less compact rocks like granite and marble, due to absorption. This is just as might have been predicted. The mean of his values of E is 9.1×10^{11} for Sudbury norite.

This compares well with the seismological value of 8.82×10^{11} dynes/cm.^{2}. Again, the very compact rock, Vinal Haven diabase, shows good agreement between the statical value of 10.2×10^{11} dynes/cm.^{2} and Dr. Muzzey's dynamical value of 9.85×10^{11} dynes/cm.^{2}.

TABLE 3

ROCK	E (MUZZEY) DYNES/CM.^{2}	E (STATICAL) DYNES/CM.^{2}
Norite, Sudbury 1	9.13×10^{11}	8.07×10^{11}
Norite, Sudbury 2	8.85	7.96
Norite, Sudbury 3	9.00	8.05
Norite, Sudbury 4	9.48	9.36
Diabase, Vinal Haven	9.85	10.2
Obsidian, Lipari	6.82	...

The discussion shows definitely that the comparison of field and statical laboratory measurements on elastic properties of rocks near the surface of the earth is a difficult matter for most rocks, because of the presence of numerous small cracks and cavities. The lack of agreement between dynamically and statically determined elastic constants is a real property of the rock, and is smaller the more compact the rock.

The cracks and cavities in rocks will doubtless cause dispersion and absorption for elastic waves. If measurements of the elastic constants E and σ are desired for comparison with data derived from seismological observations, they can be obtained by either statical or dynamical laboratory methods on highly compact rocks, and by dynamical methods *alone* on porous or loosely compacted rocks. The laboratory measurements of the compressibilities of the uncovered rocks do agree closely in every case with the seismologically measured values. Thus such laboratory measurements of the compressibility should prove of value to seismologists.

For compact surface rocks, and for rocks at depths in the earth where the pressure due to the weight above is such as to ensure freedom from the presence of cracks and cavities, it is certainly correct to treat them as homogeneous elastic media obeying Hooke's law and possessing the same statical and dynamical elastic constants.

* Paper No. 5 published under the auspices of the Committee on Geophysical Research and of the Division of Geological Sciences at Harvard University.

[1] W. A. Zisman, "Young's Modulus and Poisson's Ratio with Reference to Geophysical Applications," these PROCEEDINGS, **19**, 653–665 (1933).

[2] W. A. Zisman, "Compressibility and Anisotropy of Rocks at and near the Earth's Surface," *Ibid.*, 666–679.

[3] Leet, L. D., and Ewing, W. M., "Velocity of Elastic Waves in Granite," *Physics*. **2**, No. 3, March, 160–173 (1932).

[4] T. C. Richards, *Proc. Phys. Soc. London*, **45**, I (Jan., 1933).

THE GENERAL CASE OF NON-HOMOGENEOUS LINEAR DIFFERENTIAL EQUATIONS

BY W. J. TRJITZINSKY

DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY

Communicated June 12, 1933

1. Our present object is to establish the asymptotic nature of solutions of the equation $L_n(y) = b(x)$, where the differential operator $L_n(y)$ is of the form

$$L_n(y) \equiv a_0(x)y^{(n)}(x) + a_1(x)y^{(n-1)}(x) + \ldots + a_{n-1}(x)y^{(1)}(x) + a_n(x)y(x) \quad [a_0(x) \not\equiv 0; \quad a_n(x) \not\equiv 0] \quad (1)$$

The coefficients $a_i(x)(i = 0, 1, \ldots n)$ and $b(x)$ are analytic for $|x| \geqq p_0$ $(|x| \neq \infty)$ being representable by convergent series of the form

$$a_\mu x^{\frac{\mu}{p}} + a_{\mu-1} x^{\frac{\mu-1}{p}} + \ldots + a_1 x^{\frac{1}{p}} + a_0 + a_{-1} x^{-\frac{1}{p}} + \ldots, \quad (2)$$

where p $(\geqq 1)$ is an integer. More generally, these coefficients may be considered to be merely asymptotic, in certain regions, to series of the form (2).[1] It is assumed that not all the coefficients of the series (2), associated with $a_0(x)$ and $a_n(x)$, are zero. Thus, without any loss of generality, we take $a_0(x) \equiv 1$. *No restrictions are made concerning the roots of the characteristic equation corresponding to the homogeneous problem* $L_n(y) = 0$.

2. At the basis of the present treatment lies the asymptotic theory of the homogeneous problem, which the present author has completely developed (with no restrictions on the roots of the characteristic equation) in a detailed paper to appear elsewhere. In the sequel this paper will be referred to as (T). The development of the theory of the homogeneous equation (or linear system, as well), in (T), has been based on successive applications of a method of "Iterations" (inspired by a method so called and originally applied by G. D. Birkhoff to difference equations), factorization and suitable integrations.

Those results of (T) which are necessary for the purposes at hand will be now stated.

Let $L_n^*(y)$ denote $L_n(y)$ with the coefficients replaced by the corresponding series. As follows from a work of E. Fabry (Thèse, 1885, Paris) the formal equation $L_n^*(y) = 0$ is satisfied by a full set of formal series solutions of the form

$$s_i(x) = e^{Q_i(x)} x^{r_i} \partial_i(x), \quad Q_i(x) = \sum_{v=0}^{v=l_i-1} q_v^i x^{\frac{l_i-v}{h_i}},$$

$$\partial_i(x) = \partial_i^0(x) + \partial_i^1(x) \log x + \ldots + \partial_i^{m_i}(x) \log^{m_i}(x), \quad (3a)$$

$$\mathfrak{d}_i^{\mu}(x) = \mathfrak{d}_i^{\mu,0} + \mathfrak{d}_i^{\mu,1}x^{-\frac{1}{k_i}} + \mathfrak{d}_i^{\mu,2}x^{-\frac{2}{k_i}} + \dots \tag{3b}$$

$$(\mu = 0, 1, \dots m_i;\ i = 1, 2, \dots n),$$

where l_i, m_i, k_i are integers [$m_i \geqq 0$; $k_i = r_i p$; integer $r_i \geqq 1$].[2]

A series $\mathfrak{d}(x)$ of the form (3a), (3b) is called a $\mathfrak{d}$-series.

Branches, extending to infinity, along which

$$RQ_{i,j}(x) = 0 \quad [Q_{i,j}(x) = Q_i(x) - Q_j(x)] \tag{4}$$

for some i and j ($i \neq j$) will be termed Q curves.[3] Let k be a suitable integer. *The complete vicinity of infinity*

$$\theta_0 \leqq \theta = \arg x \leqq 2\pi k + \theta_0 \tag{5}$$

$$(|x| = \rho \geqq \rho_1 > 0)$$

is divided in a number of successively adjacent regions

$$R_1, R_2, \dots R_n \tag{5a}$$

separated by Q curves; interior neither of these regions are there any Q curves. The region R_1 will have a certain Q curve in common with R_n. Within any particular region R_i a certain ordering of the values $RQ_i(x)$ is maintained.

Let R_i denote any particular one of the regions (5a), whose boundaries have different limiting directions at infinity. Let

$$Q_{i_H j_H}(x) \equiv Q_{i_H}(x) - Q_{j_H}(x) \quad [\not\equiv 0;\ H = 1, \dots \lambda] \tag{6}$$

constitute the totality of differences $Q_{i,j}(x)$, each of which has a non-negative real part in R_i and is such that for some β

$$|x^{-\beta}|\, e^{RQ_{i_H, j_H}(x)} \longrightarrow 0 \tag{6a}$$

along both boundaries of R_i. Suppose there are some such functions (6). *Let R_i' denote then a subregion of R_i with one boundary in common with R_i and another boundary, interior R_i, such that along it all the left members of* (6a) *increase indefinitely for every β (> 0). We let R_i'' denote an analogous subregion of R_i, with the other boundary in common with R_i.* The limiting directions (at infinity) of the two boundaries of R_i' are correspondingly the same as those of R_i; the same is true of R_i'' (provided the regions are suitably specified, as is possible).

The main result of (T) is embodied in the theorem.

THE FUNDAMENTAL EXISTENCE THEOREM. *Consider an equation $L_n(y) = 0$ of the type specified above. Let R_i be any particular region of the set of regions* (5a). *If there exist no functions* (6) *such that* (6a) *holds, or if the boundaries of R_i have the same limiting directions (at infinity), there exists a full set of analytic solutions $y_v(x)$ ($v = 1, 2, \dots n$) such that*

$$y_v(x) \sim e^{Q_v(x)} x^{r_v} \partial_v(x) = s_v(x) \qquad (7)$$

$$(v = 1, 2, \ldots n; \quad x \text{ in } R_i).$$

The asymptotic relations here are with respect to the power series factors. If there exist functions (6) *such that* (6a) *holds, while the boundaries of* R_i *have different limiting directions, consider the two subregions* R_i' *and* R_i'' *of* R_i *(as specified in the italics following* (6a)). *In this case, also, there will exist a full set of solutions satisfying* (7) *in* R_i', *and another set of solutions satisfying* (7) *in* R_i''. *The* $s_i(x)$ *are formal solutions of the formal equation* $L_n^*(y) = 0$.

Analogous results hold for linear differential systems.

3. Consider now the non-homogeneous problem. Let R denote any particular region for which existence of solutions (of the equation $L_n(y) = 0$), satisfying the asymptotic relations (7), had been asserted in the Fundamental Existence Theorem. Thus, R may denote R_i (a region of the set (5a)) or it may denote one of the regions R_i', R_i''. This region R will be subdivided further as follows. Consider the branches, extending to infinity, along which

$$RQ_i(x) = 0 \qquad (8)$$

for some i. In general, of course, there are several branches satisfying a particular equation (8). R will consist of a number of consecutively adjacent regions

$$R^{(i)} \quad (i = 1, 2, \ldots h) \qquad (9)$$

separated by branches each satisfying an equation (8); moreover, no such curve is to be interior a region $R^{(i)}$. It may happen that there is only one such region $R^{(i)}$, necessarily coincident with R.

Let $R^{(i)}$ be a particular region of the set (9) with the limiting directions (at infinity) of its boundaries distinct. Let

$$-Q_{i_H}(x) \quad [\not\equiv 0; \ H = 1, 2, \ldots q) \qquad (10)$$

denote the totality of functions $-Q_j(x)$, each of which has a positive real part interior $R^{(i)}$ and is such that for some β

$$\left| x^{-\beta} \right| e^{-RQ_{i_H}(x)} \longrightarrow 0 \qquad (10a)$$

along both boundaries of $R^{(i)}$. It may occur that, given a particular region $R^{(i)}$, there exist no corresponding functions (10) such that (10a) holds. Suppose there are such functions (10). *We let then* $R^{(i)1}$ *denote a subregion of* $R^{(i)}$ *with one boundary in common with* $R^{(i)}$ *and another boundary, interior* $R^{(i)}$, *such that along it the left members in* (10a) ($H = 1, 2, \ldots q$) *increase indefinitely for every* β (> 0). *On the other hand,* $R^{(i)2}$ *will denote a similar subregion of* $R^{(i)}$, *with the other boundary in common*

with $R^{(i)}$. These regions can be so constructed that the limiting directions of their boundaries are correspondingly the same as those of $R^{(i)}$.

Let $y_i(x)$ $[\equiv y_{i,1}(x);\ i = 1, 2, \ldots n]$ constitute a fundamental set of solutions of the equation $L_n(y) = 0$, satisfying in R the asymptotic relations (7) of the fundamental Existence Theorem. Let $(y_{i,1}^{(j-i)}(x))$ denote a matrix with $y_{i,1}^{(j-1)}(x)$ in the i-th row and j-th column $(i,j = 1, 2, \ldots n)$. A solution of $L_n(y) = b(x)$ would be given by

$$y(x) = \sum_{\lambda=1}^{n} y_{\lambda,1}(x) \int b(x)\, \bar{y}_{n,\lambda}(x)dx, \tag{11}$$

where the $\bar{y}_{i,j}(x)$ are elements of the inverse of the matrix $(y_{i,1}^{(j-1)}(x))$; that is,

$$(\bar{y}_{i,j}(x)) \doteq (y_{i,1}^{(j-1)}(x))^{-1}.$$

The integrations in (11) are specified, for a region $R^{(i)}$ or $R^{(i)\prime}$ (or $R^{(i)\prime\prime}$), as the case may be, according to the integration methods used in (T). Application of these methods is possible in view of the conditions satisfied by the regions $R^{(i)}$, $R^{(i)\prime}$, $R^{(i)\prime\prime}$.

On taking account of (7) and of the asymptotic form of $b(x)$ it will follow that, for x in $R^{(i)}$ or $R^{(i)\prime}$ (or $R^{(i)\prime\prime}$),

$$y(x) \sim x^{\frac{l}{k_1}} \mathfrak{d}(x), \tag{12}$$

where l and k_1 (> 0) are integers and $\mathfrak{d}(x)$ is a $\mathfrak{d}$-series.

EXISTENCE THEOREM. *Let* $L_n(y) = b(x)$ *be an equation of the type specified in* (§ 1). *Let* $R^{(i)}$ *be any particular region of the set of regions* (9). *If there exist no functions* (10) *such that* (10a) *holds, or if the boundaries of* $R^{(i)}$ *have the same limiting directions at infinity, there exists an analytic solution* $y(x)$ *satisfying the asymptotic relation* (12) *in* $R^{(i)}$. *If there exist functions* (10) *such that* (10a) *holds, while the boundaries of* $R^{(i)}$ *have different limiting directions, we consider two subregions* $R^{(i)\prime}$ *and* $R^{(i)\prime\prime}$ *of* $R^{(i)}$ *(as specified in the italics following* (10)). *In this case, also, there exists an analytic solution* $y(x)$ *satisfying* (12) *in* $R^{(i)\prime}$, *and another solution satisfying* (12) *in* $R^{(i)\prime\prime}$. *The right member in* (12) *constitutes a formal solution of the formal non-homogeneous equation corresponding to the equation* $L_n(y) = b(x)$.

[1] Correspondingly, the results will be asserted for suitably restricted regions.

[2] Except with x and $\log x$, superscripts here do not denote powers.

[3] $RQ(x)$ denotes "real part of $Q(x)$."

A CANONICAL FORM FOR A SET OF VECTORS

BY M. S. KNEBELMAN

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated April 22, 1933

By a canonical form for a set of vectors we understand the form of the components of these vectors in a particular coördinate system in which these components have some desired simplicity. In this note we propose to obtain a canonical form for a set of contravariant and for a set of covariant vectors. Without loss of generality we may assume that the given set is independent and we shall also assume that the set is complete since the reduction of an incomplete set is accomplished by reducing the completed set.

1. Let $\xi^{\alpha}_{(i)}(x)$ be a set of r contravariant vectors in an n-dimensional space, $r \leq n$. The vectors being independent implies that the matrix $\| \xi^{\alpha}_{(i)} \|$ is of rank r. We shall suppose that the functions $\xi(x)$ are analytic in the neighborhood of the origin. The set being complete, we have by definition

$$\xi^{\gamma}_{(i)} \frac{\partial \xi^{\alpha}_{(j)}}{\partial x^{\gamma}} - \xi^{\gamma}_{(j)} \frac{\partial \xi^{\alpha}_{(i)}}{\partial x^{\gamma}} = C^{h}_{ij}(x) \xi^{\alpha}_{(h)}, \tag{1.1}$$

where $C^{h}_{ij}(x)$ are the *structure functions* of the set. The rank of the matrix $\| \xi^{\alpha}_{(i)} \|$ being r, a transformation of coördinates, which amounts to a mere renumbering, will insure the fact that $| \xi^{i}_{(j)} | \neq 0$, $i,j = 1, \ldots, r$. Then the system of differential equations

$$\xi^{\alpha}_{(i)} \frac{\partial \varphi}{\partial x^{\alpha}} = 0 \tag{1.2}$$

admits (because of (1.1)) $n - r$ independent solutions $\varphi^{\sigma}_{(x)}$, $\sigma = r + 1, \ldots, n$. When we apply to our set of vectors a transformation of coordinates defined by $\bar{x}^{\sigma} = \varphi^{\sigma}_{(x)}$, $\bar{x}^{i} = x^{i}$, in the new coördinate system[1] we shall have

$$\bar{\xi}^{\sigma}_{(i)} \equiv 0 \; \sigma > r. \tag{1.3}$$

Moreover, any transformation of the form $x'^{i} = x'^{i}(x^{1} \ldots x^{n})$ $x'^{\sigma} = x'^{\sigma}(x^{r+1}, \ldots, x^{n})$ will preserve the property (1.3). We have therefore reduced our set of vectors to a form in which all components after the r^{th} are zero and such that

$$\xi^{h}_{(j)} \frac{\partial \xi^{i}_{(k)}}{\partial x^{h}} - \xi^{h}_{(k)} \frac{\partial \xi^{i}_{(j)}}{\partial x^{h}} = C^{h}_{jk} \xi^{i}_{(h)}, \tag{1.4}$$

where the coördinates $x^{r+1}, \ldots, x^n$—if they enter at all—may be treated as parameters. A transformation of the coördinates $x^1, \ldots, x^r$ can always be found such that $\xi^i_1 = \delta^i_{(1)}$. In this coördinate system (1.4) becomes

$$\frac{\partial \xi^i_{(k)}}{\partial x^1} = C^h_{1k}\, \xi^i_{(h)}, \tag{1.5}$$

$$\xi^h_{(j_1)} \frac{\partial \xi^i_{(k_1)}}{\partial x^h} - \xi^h_{(k_1)} \frac{\partial \xi^i_{(j_1)}}{\partial x^h} = C^h_{j_1 k_1}\, \xi^i_{(h)},\ j_1,\ k_1 > 1.$$

From the first of these it follows that

$$\xi^i_{(k)}(x^1 \ldots x^n) = \xi^i_h\ (0, x^2, \ldots, x^n) \cdot A^h_{1k}(x^1 \ldots x^n), \tag{1.6}$$

where A^h_{1k} is a non-singular matrix defined by

$$A^h_{1k} = \delta^h_k + C^h_{1k}\,(0, x^2, \ldots, x^n)x^1 + C^h_{1k,1}\,\frac{(x^1)^2}{2!} + C^h_{1k,1,1}\,\frac{(x^1)^3}{3!} + \ldots \tag{1.7}$$

a subscript preceded by a comma indicating *intrinsic covariant differentiation*, i.e., if F^h is a set of scalar functions then

$$F^h_{,j} = \xi^i_{(j)}\,\frac{\partial F^h}{\partial x^i} + F^k C^h_{jk}. \tag{1.8}$$

It can be easily verified that

$$\frac{\partial A^h_{1k}}{\partial x^1} = C^m_{1k}\ A^h_{1m} \tag{1.9}$$

and by means of this to establish that, if $\overline{A}_1$ is the inverse matrix of A_1, then

$$\frac{\partial \overline{A}^k_{1l}}{\partial x^1} = C^m_{l1}\ \overline{A}^k_{1m}, \tag{1.10}$$

$$\overline{A}^k_{1l} = \delta^k_l + C^k_{l1}(0, x^2, \ldots, x^n)\ x^1 + C^k_{l1-1}\,\frac{(x^1)^2}{2!} + \ldots \tag{1.11}$$

where

$$F^h_{,j} = \xi^i_{(j)}\,\frac{\partial F^h}{\partial x^i} + F^k C^h_{kj}. \tag{1.12}$$

When (1.6) are substituted into (1.5) we get

$$\xi^h_{(\lambda)}\,\frac{\partial \xi^i_{(\mu)}}{\partial x^h} - \xi^h_{(\mu)}\,\frac{\partial \xi^i_{(\lambda)}}{\partial x^h} = \Bigg[C^h_{jk} A^P_{1k} \xi^i_{(p)} - \xi^h_{(p)} \xi^i_{(q)} \left(A^p_{1j}\,\frac{\partial A^q_{1k}}{\partial x^h} - A^p_{1k}\,\frac{\partial A^q_{1j}}{\partial x^h} \right) \Bigg] \overline{A}^j_{1\lambda}\ \overline{A}^k_{1\mu}, \tag{1.13}$$

where $\xi^h_{(\lambda)}$ are functions of $x^2, \ldots, x^n$ at most. Since the left-hand side of (1.13) is free of x^1, so is the right-hand side; therefore its value is unaltered if we put $x^1 = 0$. Now from (1.7) and (1.11) we have $A^h_{1k} = \bar{A}^h_{1k} = \delta^h_k$ and $\partial A^h_{1k}/\partial x^j = C^h_{1k}(0, x^2, \ldots)\ \delta^1_j$ for $x^1 = 0$. Thus equations (1.13) become

$$\xi^{h_1}_{(j_1)} \frac{\partial \xi^i_{(k_1)}}{\partial x^{h_1}} - \xi^{h_1}_{(k_1)} \frac{\partial \xi^i_{(j)}}{\partial x^{h_1}} = (C^h_{j_1 k_1} + C^h_{k_1 1}\xi^1_{(j_1)} + C^h_{1 j_1}\xi^1_{(k_1)})\xi^i_{(h)}, \quad (1.14)$$

where all functions involved are the values of the corresponding original functions for $x^1 = 0$.

If we consider (1.14) for $i > 1$ we shall have (because $\xi^i_{(1)} = 0$ if $i > 1$) a complete set of $r - 1$ independent vectors since $\left|\xi^{i_1}_{(j_1)}\right| = \left|\xi^i_{(j)}\right| \neq 0$, whose structure functions[2] are

$$C'^{h_1}_{j_1 k_1}(x^2, \ldots, x^n) = C^{h_1}_{j_1 k_1}(0, x^2, \ldots, x^n) + C^{h_1}_{k_1 1}\xi^1_{(j_1)} + C^{h_1}_{1 j_1}\xi^1_{(k_1)}. \quad (1.15)$$

We may now apply a transformation of coördinates of the form $\bar{x}^1 = x^1$, $\bar{x}^{i_1} = x^{i_1}(x^2, \ldots, x^n)$, $\bar{x}^\sigma = x^\sigma$, such that $\bar{\xi}^i_{(1)} = \delta^i_1 \bar{\xi}^1_{(i_1)} = \xi^1_{(i_1)}\ \bar{\xi}^{i_1}_{(2)} = \delta^{i_1}_2$ and as before we shall find from the equations of completeness that

$$\xi^{i_1}_{(j_1)} = \xi^{i_1}_{(h_1)}(0, 0, x^3 \ldots)\ A^{h_1}_{2 j_1}(x^2 \ldots x^n), \quad (1.16)$$

where

$$A^{h_1}_{2 j_1} = \delta^{h_1}_{j_1} + C'^{h_1}_{2 j_1}(0, x^3, \ldots, x^n)x^2 + C'^{h_1}_{2 j_1} 2 \frac{(x^2)^2}{2!} + \ldots$$

Obviously this process may be repeated $r - 1$ times, since for one vector there are no equations for completeness. We now observe that $\xi^i_{(1)} = \delta^i_1$ and because of (1.6) and (1.16) the second vector is such that *for* $x^1 = 0$ its components are $\xi^1_{(2)}(x^2 \ldots x^n)$, $\delta^i_2\ i > 1$. Similarly the third vector is such that *for* $x^1 = x^2 = 0$ its components are $\xi^1_{(3)}(x^3, \ldots)$, $\xi^2_{(3)}(x^3, \ldots)$, $\delta^i_3\ i > 2$ and so on. We may therefore combine these results into the following theorem:

A set of s ($\leq n$) independent contravariant vectors in an n-dimensional space being given such that the completed set contains r ($s \leq r \leq n$) independent vectors, a coördinate system may be found in which all components after the r^{th} are zero throughout the space, the remaining components being

ξ_1 : δ^i_1 *throughout the space*

$\xi_{(2)}$: $\xi^1_{(2)}$, $\delta^i_2\ i > 1$ *for* $x^1 = 0$

$\xi_{(3)}$: $\xi^1_{(3)}$, $\xi^2_{(3)}$, $\delta^i_3\ i > 2$ *for* $x^1 = x^2 = 0$

$\xi_{(s)}$: $\xi^1_{(s)}, \ldots, \xi^{s-1}_{(s)}$, $\delta^i_s\ i > s - 1$ *for* $x^1 = x^2 = \ldots x^{s-1} = 0$.

It may be shown further that the transformations involved in the above process have enough arbitrariness in them to enable one to choose the subspace V_{n-s+1} arbitrarily—except for certain restrictions—on which the s vectors are to have the above canonical form.

In particular an ennuple of vectors and a curve being given, the curve not belonging to the congruence determined by the ennuple, a coördinate system exists in which the curve has the equations $x^1 = x^2 = \ldots = x^{n-1} = 0$, and the vectors have components such that $\xi^\alpha_{(\beta)} = \delta^\alpha_\beta$ for $\alpha \geqslant \beta$.

2. A set of r ($\leq n$) independent covariant vectors may be reduced to canonical form by a similar process although there are some essential differences. Let $\eta^i_{(\alpha)}(x)$ be a set of independent vectors and without loss of generality we may assume that $|\eta^{(i)}_j| \neq 0$. The set being complete, we have by definition

$$\frac{\partial \eta^{(i)}_\alpha}{\partial x^\beta} - \frac{\partial \eta^{(i)}_\beta}{\partial x^\alpha} = C^i_{jk}(x)\, \eta^{(j)}_\alpha \eta^{(k)}_\beta, \tag{2.1}$$

the C's being the structure functions of the set. It can be shown that the system of partial differential equations $\eta^{(i)}_\alpha \partial x^\alpha / \partial \bar{x}^\sigma = 0$, $\sigma > r$ is completely integrable[3] and that a coördinate system therefore exists in which $\eta^{(i)}_\sigma \equiv 0\ \sigma > r$. Moreover, a transformation of the form

$$x^i = x^i(\bar{x}^1, \ldots, \bar{x}^r) \qquad i \leq r$$
$$x^\sigma = x^\sigma\ (\bar{x}^1, \ldots, \bar{x}^n) \quad \sigma > r$$

will preserve this characteristic of each covariant vector. In the coordinate system determined above (2.1), being tensor equations, give for $\alpha \leq r\ \beta > r$

$$\frac{\partial \eta^{(i)}_j}{\partial x^\sigma} = 0 \qquad \frac{\partial \eta^{(i)}_j}{\partial x^k} - \frac{\partial \eta^{(i)}_k}{\partial x^j} = C^i_{hl} \eta^{(h)}_j \eta^{(l)}_k. \tag{2.2}$$

From the first of these it follows that the components are free of x^{r+1}, $\ldots$, x^n and from the second—since $|\eta^{(h)}_j| \neq 0$—that C^i_{hl} are free of the same variables. Hence

A complete set of r ($\leq n$) independent covariant vectors being given, there exists a coördinate system in which the components $\eta^{(i)}_\sigma = 0\ \sigma > r$ and the remaining components are functions of $x^1, \ldots, x^r$ at most.

It may be observed that this theorem does not hold, in general, for a set of contravariant vectors.

We now look for a coördinate system in which the same single component of every vector except one shall be zero. Since $|\eta^{(i)}_j| \neq 0$ not every $(r-1)$-rowed minor is zero and without loss of generality we assume it to be the cofactor of $\eta^{(1)}_1$. Then the equations

$$\eta_j^{(i)} \frac{\partial x^j}{\partial \bar{x}^1} = \delta_1^i$$

admit a solution. Hence there exists a coördinate system in which $\eta_1^{(i)} = \delta_1^i$ and when this is put into (2.2) for $k = 1$ we get

$$\frac{\partial \eta_j^{(i)}}{\partial x^1} = C_{h1}^i \eta_j^{(h)}. \tag{2.3}$$

Hence, as in §1,

$$\eta_j^i(x^1, \ldots, x^r) = \eta_j^{(h)}(0, x^2, \ldots, x^r) B_{1h}^i(x^1, \ldots, x^r), \tag{2.4}$$

where

$$B_{1h}^i = \delta_h^i + C_{h1}^i (0, x^2, \ldots, x^r)x^1 + C_{h1;1}^i \frac{(x^1)^2}{2!} + \ldots \tag{2.5}$$

$$C_{h1;1}^i = -C_{h1\cdot 1}^i \text{ (cf. (1.12))}$$

$$\frac{\partial B_{1h}^i}{\partial x^1} = C_{k1}^i B_{1h}^k. \tag{2.6}$$

When (2.4) are substituted into (2.2) we get

$$\frac{\partial \eta_j^{(m)}}{\partial x^k} - \frac{\partial \eta_k^{(m)}}{\partial x^j} = C_{hl}^i B_{1a}^h B_{1c}^l \bar{B}_{1i}^m \eta_j^{(a)} \eta_k^{(c)} - \bar{B}_{1i}^m \left(\eta_j^{(h)} \frac{\partial B_{1h}^i}{\partial x^k} - \eta_k^{(h)} \frac{\partial B_{1h}^i}{\partial x^j} \right), \tag{2.7}$$

where $\eta_j^{(m)} = \eta_j^{(m)}(0, x^2, \ldots, x^r)$ and $\bar{B}_{1i}^m B_{1m}^j = \delta_i^j$. When i or $j = 1$ (2.7) are identically satisfied and since the left side of (2.7) is free of x^1 so is the right. Setting $x^1 = 0$ we find

$$\frac{\partial \eta_{j_1}^{(m)}}{\partial x^{k_1}} - \frac{\partial \eta_{k_1}^{(m)}}{\partial x^{j_1}} = C_{hl}^m(0, x^2, \ldots, x^r) \eta_{j_1}^{(h)} \eta_{k_1}^{(l)} \; j_1, k_1 > 1. \tag{2.8}$$

We observe that $\eta_{j_1}^{(m)}$ is a set of r covariant vectors in a space of $r - 1$ dimensions. Also since $|\eta_{j_1}^{(i_1)}| \neq 0$, the first vector is linearly dependent on the last $r - 1$ and therefore

$$\eta_{j_1}^{(1)} + \varphi_{k_1} \eta_{j_1}^{(m_1)} = 0, \tag{2.9}$$

the φ's being determined by these equations. Hence (2.8) for $m > 1$ reduce to

$$\frac{\partial \eta_{j_1}^{(i_1)}}{\partial x^{k_1}} - \frac{\partial \eta_{k_1}^{(i_1)}}{\partial x^{j_1}} = C'^{i_1}_{h_1 l_1} \eta_{j_1}^{(h_1)} \eta_{k_1}^{(l_1)}, \tag{2.10}$$

where

$$C'^{i_1}_{h_1 l_1} = C^{i_1}_{h_1 l_1}(0, x^2, \ldots, x^r) + C^{i_1}_{l_1 1} \varphi_{h_1} + C^{i_1}_{1 h_1} \varphi_{l_1}. \tag{2.11}$$

We are thus confronted with the same problem we had at the beginning of this section except that the number of vectors and of variables is diminished by one. As before we can find a coördinate system in which (2.12)

$$\eta_2^{(i_1)} = \delta_2^{i_1} \tag{2.12}$$

and because of (2.10)

$$\eta_{j_1}^{(i_1)} = \eta_{j_1}^{h_1}(0, x^3 \ldots x^r)\, B_{2h_1}^{i_1}(x^2 \ldots x^r), \tag{2.13}$$

where $B_{2h_1}^{i_1}$ are defined analogously to B_{1h}^{i} (cf. (2.5)). Hence $\eta_1^{(i)} = \delta_1^i$ throughout the space and *for* $x^1 = 0$ $\eta_2^{(1)} = \eta_2^{(1)}(x^2, \ldots, x^r)$ $\eta_2^{(i_1)} = \delta_2^{i_1}$ $(i_1 > 1)$. Similarly the third components of the vectors are such that for $x^1 = x^2 = 0$ they reduce to $\eta_3^{(1)}(x^3 \ldots)$, $\eta_3^{(2)}(x^3 \ldots)$ and $\delta_3^{i_2}$ $(i_2 > 2)$. Hence

A complete set of r $(\leq n)$ *independent covariant vectors* $\eta_\alpha^{(i)}(x^1, \ldots, x^n)$ *being given, a coördinate system can be found in which* $\eta_\sigma^{(i)} = 0$ *for* $\sigma > r$, *the first components are* δ_1^i *throughout the space, the second reduce to* $\eta_2^{(1)}(x^2, \ldots, x^r)$, $\delta_2^{i_1}$ $i_1 > 1$ *for* $x^1 = 0$, *the third components reduce to* $\eta_3^{(1)}$, $\eta_3^{(2)}$, $\delta_3^{i_2}$ $i_2 > 2$ *for* $x^1 = x^2 = 0$, *etc., the* r^{th} *components reducing to* $\eta_r^{(1)}$, (x^r), $\ldots$ $\eta_r^{r-1}(x\eta)$, 1 *on the subspace* V_{n-r+1} *defined by* $x^1 = x^2 = \ldots = x^{r-1} = 0$.

[1] That this transformation is non-singular is easily proved.

[2] The structure functions of a complete set of vectors besides being skew-symmetric satisfy Jacobi's identities which may be written as $C_{ij,k}^h + C_{jk,i}^h + C_{ki,j}^h = 0$. It is easily verifiable that $C_{j_1k_1}^{ih_1}$ satisfy identities of the same sort.

[3] This system is of a type similar to that studied by König and its general solution depends on n arbitrary functions of $n - r$ variables thus insuring the non-singularity of the transformation of coördinates.

ON A THEOREM OF FEJÉR

BY W. C. RANDELS

DEPARTMENT OF MATHEMATICS, BROWN UNIVERSITY

Communicated May 25, 1933

L. Fejér has shown[1] that, if a function $f(x)$ is odd, and positive and convex upward in the interval $(0, \pi)$, the $(C, 3)$ transforms of the partial sums of its Fourier series will be convex upward in the same interval, and, if the function is symmetric about the point $x = \pi/2$, the $(C, 2)$ transforms will also be convex in that interval. In this paper we propose to show that in both cases the indices obtained by Fejér are the lowest possible.

Let us define the function

$$f(x) = \begin{cases} \dfrac{x}{b}, & 0 \leq x \leq b \\ \dfrac{\pi - x}{\pi - b}, & b < x \leq \pi, \end{cases} \qquad f(-x) = -f(x).$$

Then

$$f(x) = \sum_{n=1}^{\infty} \frac{2}{b(\pi - b)n^2} \sin nb \sin nx.$$

The nth (C, α) transform of the Fourier series of $f(x)$ is given by

$$\frac{2S_n^{\alpha}(x)}{b(\pi - b)\binom{n+\alpha}{n}},$$

where

$$S_n^{\alpha}(x) = \sum_{\nu=1}^{n} \binom{n-\nu+\alpha}{n-\nu} \frac{\sin \nu b \sin \nu x}{\nu^2}.$$

The factor multiplying S_n^{α} is positive so we may disregard it and determine the condition that S_n^{α} be convex upward. This condition requires that $D_x^2 S_n^{\alpha}(x)$ must be ≤ 0. But $D_x^2 S_n^{\alpha}(0) = 0$, and therefore we must have $D_x^3 S_n^{\alpha}(0) \leq 0$. Differentiating we get

$$t_n^{\alpha} = -D_x^3 S_n^{\alpha}(0) = \sum_{\nu=1}^{n} \binom{n-\nu+\alpha}{n-\nu} \nu \sin \nu b.$$

We now investigate

$$t_2^{\alpha} = (\alpha + 1) \sin b + 2 \sin 2b = (\alpha + 1) \sin b + 4 \sin b \cos b.$$

If this is to be non-negative we must have $\alpha \geqslant -4 \cos b - 1$. Now let $b \longrightarrow \pi$, then the expression on the right tends to 3 and hence we must have $\alpha \geqslant 3$.

If $b = \pi/2, f(x)$ is clearly symmetric about the point $x = \pi/2$. It is readily found that

$$t_n^0 = \sum_{\nu=1}^{n} \nu \sin \nu \frac{\pi}{2} = (-1)^{\left[\frac{n-1}{2}\right]} \left[\frac{n+1}{2}\right].$$

Also

$$t_n^1 = \sum_{\nu=1}^{n} t_\nu^0 = 1, \quad \frac{n}{2} + 1, 0, -\frac{n}{2},$$

according as $n \equiv 1, 2, 3, 0 \pmod 4$.
If $\beta > \alpha$, it is well known that

$$S_n^\beta(x) = \sum_{\nu=1}^{n} \binom{n-\nu+\beta-\alpha-1}{n-\nu} S_\nu^\alpha(x)$$

and

$$\binom{n-\nu+\beta-\alpha-1}{n-\nu} = \frac{(n-\nu+\beta-\alpha-1)\ldots(\beta-\alpha)}{(n-\nu)!} > 0.$$

A linear combination, with positive coefficients, of functions which are convex upward is itself convex upward, hence if $S_n^\alpha(x)$ is convex upward for all n, $S_n^\beta(x)$ will also be convex upward for all n. Therefore, since $S_n^1(x)$ is clearly not convex upward for all n, we must have $\alpha > 1$. We will now suppose $1 < \alpha < 2$. Then

$$t_n^\alpha = \sum_{\nu=1}^{n} \binom{n-\nu+\alpha-2}{n-\nu} t_\nu^1.$$

Let $n = 4m$, then

$$t_n^\alpha = \sum_{\nu=1}^{n}{}' \binom{n-\nu+\alpha-2}{n-\nu} + \sum_{\nu=1}^{n}{}'' \frac{\nu}{2} \left\{\binom{n-\nu+2+\alpha-2}{n-\nu+2} - \binom{n-\nu+\alpha-2}{n-\nu}\right\},$$

where the first sum is taken over integers of the form $\nu \equiv 1 \pmod 4$ and the second over integers of the form $\nu \equiv 0 \pmod 4$. But

$$\binom{n-\nu+2+\alpha-2}{n-\nu+2} - \binom{n-\nu+\alpha-2}{n-\nu} = \frac{(n-\nu+\alpha)\ldots(\alpha-1)}{(n-\nu+2)!}$$

$$- \frac{(n-\nu+\alpha-2)\ldots(\alpha-1)}{(n-\nu)!} = \{(n-\nu+\alpha)(n-\nu+\alpha-1)$$

$$- (n-\nu+2)(n-\nu+1)\} \frac{(n-\nu+\alpha-2)\ldots(\alpha-1)}{(n-\nu+2)!} < 0$$

and the last term in the second sum is

$$\frac{n}{4}(\alpha-2)(\alpha+1) < 0.$$

Also

$$\sum_{\nu=1}^{n}{}' \binom{n-\nu+\alpha-2}{n-\nu} < \sum_{\nu=0}^{n} \binom{n-\nu+\alpha-2}{n-\nu} = \binom{n+\alpha-1}{n} = 0\,(n^{\alpha-1}).$$

This shows that t^{α}_{4m} will be negative for sufficiently large values of m, if $\alpha < 2$.

[1] In the address delivered at the Chicago meeting of the American Mathematical Society, which is forthcoming in the *Bull. Amer. Math. Soc.*

---

*ARISTOGENESIS, THE OBSERVED ORDER OF BIOMECHANICAL EVOLUTION**

BY HENRY FAIRFIELD OSBORN

THE AMERICAN MUSEUM OF NATURAL HISTORY

Read before the Academy, Tuesday, April 25, 1933

The term "biomechanical"[1] applies to organic as distinguished from inorganic processes;[2] for example, the evolution of the living invertebrates and vertebrates of the past as observed in paleontology is chiefly recorded in the hard parts of animals formerly full of life. For the observed order of biomechanical adaptation and evolution, which is the subject of the present communication, no fitter, no more apt term can be found than *aristogenesis* proposed but not defined in my contribution to the "Centenary of Evolution" symposium of the British Association for the Advancement of Science in 1931 and in my last contribution to the National Academy.[3]

First, the term aristogenesis is derived from the Attic Greek "aristos" (αριστος) signifying "*best in its kind*," in Euripides referring to warriors, the "*best*," the "*most useful*;" in Homer referring to chariots, the "*best*" or "*finest*."[4] The Greek "genesis" (γενεσις) from the Aryan base meaning "*to produce, bring into being*" means "*origin, source, birth, descent; . . . mode of formation or production;*"[5] also, in Plato, "*an origin, source, productive cause; origination, making (even of common things).*"[6]

Second, the term *aristogenesis* relates to an actually observed order of nature, a concrete process; it has no theoretic significance such as the "entelechy" of Aristotle, the "pangenesis" of Darwin, the "vitalism" of

* This is the ninth contribution on the Origin of Species and the principles of biomechanical evolution as demonstrated in paleontology. The numerical references in the present communication are from the complete bibliography in Osborn's "Fifty-Two Years of Research" which will be supplied on application to the Osborn Library.

Driesch, the "bathmism" of Cope, the "emergence" of Morgan, the "holism" of Smuts. It is rather to be compared with the "biogenesis" of Haeckel, a term signifying the recapitulation of ancestral evolution in embryonic and foetal development.

Third, aristogenesis as a process of nature is not to be challenged as a concept, still less as an explanation, hypothesis or theory. At present we know how aristogenesis works, we have set forth (Osborn: 874, 901) a series of principles of its operation, but we have no explanation or hypothesis to offer as to why it works or what are its causes. In this sense aristogenesis can only be challenged by a reëxamination and reinterpretation of the paleontological evidence on which it rests. One principle of this evidence was discovered in the invertebrates by Waagen in 1869, namely, the orderly "W. mutations"[7] of the shells of molluscs which arise in a most inconspicuous manner and develop into conspicuous organs. New and amplified principles have been discovered in vertebrates as fully enumerated in my last two communications (Osborn: 874, 901) in this series of biological inductions from my own paleontological researches of the last half century. (See full bibliographic list below.)

The evidence for aristogenesis or the direct origin from the geneplasm of biomechanical adaptations is now overwhelming; Waagen's principle of W. mutations in invertebrates (see Osborn: 353) of 1869 was re-discovered in 1894 by Osborn (Osborn: 92) in the teeth of vertebrates. It has been confirmed and amplified by researches recorded in fifty-one contributions between the years 1894 and 1933, chiefly based upon thousands of independent observations on the evolution of the grinding teeth (Osborn: 33) in the primates, in the equines and proboscideans (teeth), rhinoceroses (teeth and horns), and titanotheres (teeth and horns). These observations are harmonious and concurrent. They have been made purely inductively with no other purpose than the direct recording of facts. Now they enable us to give a clear definition of aristogenesis as follows:

Aristogenesis is the gradual, secular, continuous, direct, reactive, adaptive origin of new biomechanisms. It is a creative process from the geneplasm of entirely new germinal characters. It is the orderly creation of something better or more adaptive. Certain lines of descent are distinguished by the potentiality of creative origin from the geneplasm of new adaptive biomechanisms. Germinally predetermined origin of new characters tends toward betterment, arises independently, in widely separated geographic areas, at the same or different aristogenic rates. Aristogenesis is a secular genetic and adaptive reaction rather than an immediate adaptive reaction to new habit or environment as in Lamarck's theory. Aristogenes are independent in origin and early development of the Natural Selection principle of Darwin.

Fourth, since single new biomechanisms originate in the geneplasm (a term preferable to germ plasm) they may be called "aristogenes" as

distinguished from "somatogenes" which originate in the somaplasm. Aristogene is also a better and more distinctive term than "rectigradation," the term first applied (Osborn: 314) to these new biomechanisms.

Fifth, aristogenesis, as the direct adaptive origin of new biomechanisms, is of two kinds:

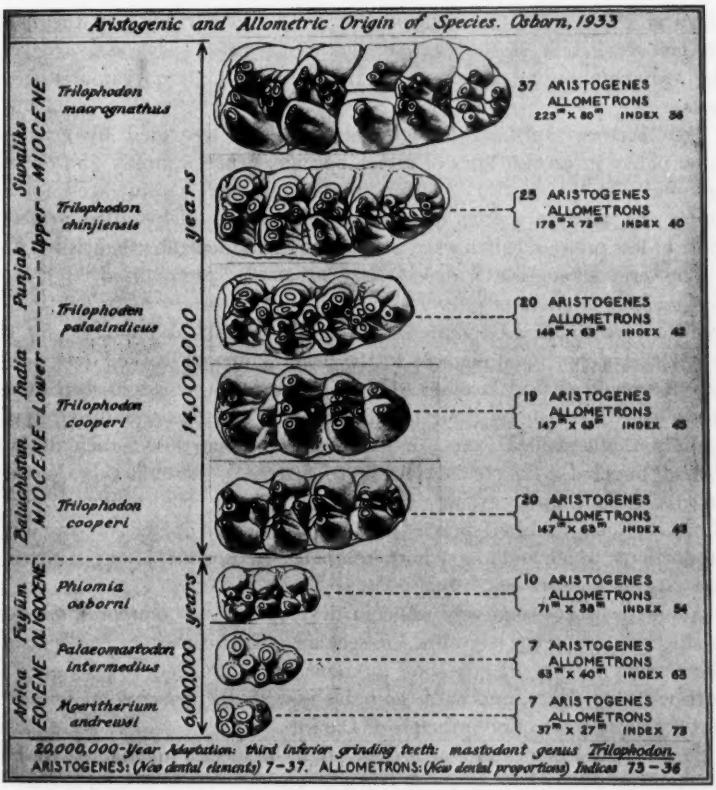

Aristogenic and Allometric Origin of Species. Osborn, 1933
Trilophodon macrognathus
Trilophodon chinjiensis
Trilophodon palaeindicus
Trilophodon cooperi
Trilophodon cooperi
Phiomia osborni
Palaeomastodon intermedius
Moeritherium andrewsi
Siwaliks
Punjab
India
Baluchistan
MIOCENE-Lower - - - - - - - - - - Upper - MIOCENE
Fayûm
OLIGOCENE
EOCENE
14,000,000 years
6,000,000 years
37 ARISTOGENES ALLOMETRONS 225m x 80m INDEX 36
25 ARISTOGENES ALLOMETRONS 178m x 72m INDEX 40
20 ARISTOGENES ALLOMETRONS 148m x 63m INDEX 42
19 ARISTOGENES ALLOMETRONS 147m x 63m INDEX 43
20 ARISTOGENES ALLOMETRONS 147m x 63m INDEX 43
10 ARISTOGENES ALLOMETRONS 71m x 38m INDEX 54
7 ARISTOGENES ALLOMETRONS 63m x 40m INDEX 63
7 ARISTOGENES ALLOMETRONS 37m x 27m INDEX 73
20,000,000-Year Adaptation: third inferior grinding teeth: mastodont genus Trilophodon.
ARISTOGENES: (New dental elements) 7-37. ALLOMETRONS: (New dental proportions) Indices 73-36

FIGURE 1

Aristogenic evolution from *Phiomia osborni* to *Trilophodon macrognathus* of the third inferior grinding tooth in the long-jawed mastodons migrating from northern Africa (*Phiomia*) to southern India (*Trilophodon*).

(*a*) ARISTOGENES (rectigradations) are governed or predetermined by germinal potentiality in certain lines of racial, specific, generic, family and ordinal descent. The closer the descent or phyletic kinship the more similar are the aristogenes. In extremely close racial lines like that of the

fourteen-million year longirostrine mastodons cited in my last communication to the Academy (Osborn: 901), the aristogenes are closely similar no matter in what part of the world they arise. Thus from 10 aristogenes in the lower Oligocene mastodon (*Phiomia osborni* of the Fayûm of northern Egypt) the number rises to 19 in *Trilophodon cooperi* also in *Trilophodon pontileviensis* of the basal Miocene of France and of India, to 20 in the lower Miocene *Trilophodon palæindicus*, to 25 in the *Trilophodon chinjiensis* of the Punjab and 37 in *Trilophodon macrognathus*, the two culminating members of this fourteen-million year aristogenic series.

(*b*) ALLOMETRONS are not governed or predetermined by germinal potentiality in certain lines of racial, specific, generic, family and ordinal descent. On the contrary, within species and even within races, for example the modern species of man *Homo sapiens*, diverse allometrons or more or less profound changes of proportion, are independently arising.

The term allometron (Greek ἀλλοιος *"of another sort or kind, different"*[8] and μετρον *"to measure in any way"*[9]) signifies changes of proportion or intensity which may be expressed in measurements and indices, for example the proportional changes of the head expressed in the Greek terms brachycephaly (broad headed), dolichocephaly (long headed), brachyopy (short faced), dolichopy (long faced), bathycephaly (deep skulled), hypsicephaly (high skulled), etc., or in the limb proportions brachydactyly (broad fingered), dolichodactyly (long fingered), leptodactyly (slender fingered), dolichopody (long footed), brachiopody (short footed), dolichomely (long limbed), brachymely (short limbed), etc., or in tooth proportions brachyodonty (short toothed), hypsodonty (long toothed), etc.

Allometrons are relatively rapid in development, or temporal, whereas aristogenes are relatively slow, or secular. The causes of geneplasmic allometrons are obscure. In some unknown way they seem to be connected both with hormones, and with the habit inheritance principle of Lamarck and the natural selection principle of Darwin.

So far as the modes in which the "origin of species" is defined by biomechanisms we are now on absolutely sure ground. This ground is contra-Lamarckian and contra-Darwinian. It is also contrary to the neo-Darwinian evolutionary hypotheses of the leading biologists and geneticists of our day.

BIBLIOGRAPHIC LIST OF THE ORIGIN OF SPECIES SERIES OF PAPERS BY HENRY FAIRFIELD OSBORN

1. 1925: 649. "The Origin of Species as Revealed by Vertebrate Paleontology," *Science*, N. S., **61**, 23, 24.

1925; 654. "The Origin of Species as Revealed by Vertebrate Paleontology," *Nature*, **115**, 925, 926; 961-963.

2. 1925: 672. "The Origin of Species, II—Distinctions between Rectigradations and Allometrons," *Proc. Nat. Acad. Sci.*, **11**, 749–752.

3. 1926: 681. "The Origin of Species, 1859–1925," *Bull. 1, Peabody Museum of Nat. Hist.*, 25–38; *Scientific Monthly*, **22**, March, pp. 185–192.

4. 1926: 693. "The Problem of the Origin of Species as It Appeared to Darwin in 1859 and as It Appears to Us Today," *Nature*, **118**, pp. 270–273; *Science*, N. S., **64**, pp. 337–341; (Abstr.) *Brit. Assoc. Adv. Sci., Rept. of the Ninety-Fourth Meeting*, 1926, p. 359.

4A. 1926: 701. "The Problem of the Origin of Species as It Appeared to Darwin in 1859 and as It Appears Today (a Correction)," *Nature*, **118**, 591, 592.

5. 1927: 714. "The Origin of Species, V—Speciation and Mutation," *Amer. Nat.*, **61**, pp. 5–42.

6. 1932: 874. "The Nine Principles of Evolution Revealed by Paleontology," *Amer. Naturalist*, **66**, Jan.-Feb., pp. 52–60.

7. 1931: 870. "New Concept of Evolution Based upon Researches on the Titanotheres and the Proboscideans," *Science*, N. S., **74**, No. 1927, pp. 557–559.

8. 1932: 897. "Biological Inductions from the Evolution of the Proboscidea," *Science*, Dec. 2, 1932, **76**, No. 1979, pp. 501–504.

1933: 901. "Biological Inductions from the Evolution of the Proboscidea," *Proc. Nat. Acad. Sci.*, **19**, No. 1, pp. 159–163, January, 1933.

9. 1933: 906. "Aristogenesis, the Oberved Order of Biomechanical Evolution," *Nature* (in press). (Abstract, with same title, *Science*, **77**, No. 2004, 519.)

10. 1933. "Aristogenesis, the New Inductive Principle of Biomechanical Evolution" (address delivered before the Académie des Sciences, Paris, May 29, 1933).

[1] Not found in "The Universal Dictionary of the English Language," edited by Henry Cecil Wyld, London, 1932; or the "Dictionary of Philosophy and Psychology," edited by James Mark Baldwin, New York and London (1901).

[2] Osborn, Henry Fairfield, "Biological Inductions from the Evolution of the Proboscidea," *Proc. Nat. Acad. Sci.*, **19**, No. 1, 159–163 (1933).

[3] *Idem;* also, "The Nine Principles of Evolution Revealed by Paleontology," *The Amer. Nat.*, **66**, Jan.-Feb., 52–60 (1932).

[4] "A Greek-English Lexicon," compiled by Henry George Liddell and Robert Scott (7th ed.), New York, p. 218 (1883).

[5] Wyld, op. cit., 476–477.

[6] Liddell and Scott, op. cit., p. 304.

[7] As distinguished from the fortuitous "D. mutations" of De Vries.

[8] "A Greek-English Lexicon," Liddell and Scott, p. 65.

[9] *Ibid.*, p. 956.

ELECTROSMOSIS AND ANOMALOUS OSMOSIS

BY DAVID R. BRIGGS

OTHO S. A. SPRAGUE MEMORIAL INSTITUTE AND THE DEPARTMENT OF PATHOLOGY, UNIVERSITY OF CHICAGO

Communicated June 12, 1933

That electrosmosis can account for observed anomalies in osmotic transfer of liquid across a membrane has been held by Girard,[1] Loeb[2] and Bartell.[3] Their measurements have been made upon rates and direction of flow of fluid through various membranes bathed on the two sides by electrolyte solutions of different concentrations. Since, in such cases, the membranes were slowly permeable to the electrolytes used, equilibrium measurements were not possible. These dynamic measurements, while indicating the magnitude of the resultant of the forces acting upon the fluid within the membrane at the time of observation, were not of such a nature that the magnitude of these forces could be individually evaluated.

Certain prerequisites to the establishing of an electrosmotic pressure across a membrane are necessary. There must exist a difference in electrical potential across the membrane, there must be an electrokinetic potential existing between the wall of the pores of the membrane and the fluid filling the pore and a flow of current must take place as a result of which the charged, movable layer of liquid in the pore is moved, carrying along with it the liquid in the center of the pore.

Numerous measurements have been made upon the osmotic pressure exhibited by solutions containing colloidal electrolytes. In such cases, the membrane is impermeable to the colloid ion and the diffusible ions in the system arrange themselves at equilibrium in a manner described in Donnan's[4] membrane equilibrium theory. Loeb[5] found that the osmotic pressure shown by casein in HCl solution was entirely due to the unequal distribution of these diffusible ions, i.e., the colloidal micellae exerted no detectable osmotic pressure. Bjerrum[6] studied a chromium hydroxide solution, and Adair[7] studied hæmoglobin. Both concluded that, within certain ranges of salt, hydrion and colloid concentration the observed osmotic pressures were the sum of the pressures derived from the unequal distribution of diffusible ions and that derived from the colloid micellae. Hammarsten,[8] using sodium nucleate, and Samac,[9] studying amylopectin and ligno sulphuric acid, found that the observed osmotic pressure was less than that calculated on the basis of differences in diffusible ion concentration and concluded that interionic forces, exerted by the large colloid ion, completely or to a large extent destroyed the osmotic acitivity of the small ions (the *Gegenionen*) present (and ionized from the colloid). With such varied conclusions it seems probable that some unrecognized force may have been acting in most cases to modify the true osmotic pressures.

In all such experiments the conditions necessary to the establishment of an electrosmotic pressure were present. The membrane potential resulting from unequal distribution of diffusible ions on the two sides of the membrane could function as the electrical potential across the membrane needed to cause liquid to be transferred electrosmotically through the pores where, undoubtedly, there must exist a ζ-potential of one sign or the other. The circuit through which the current must flow in order that the water may be moved electrosmotically through the pore could be completed within the pore, i.e., in one direction the current would be carried by the ions making up the movable half of the double layer along the pore wall and in the other direction by ions within the body of the liquid in the pore (see Bartell,[3] Freundlich[10]). At equilibrium, a pressure would be set up as a result of the electrosmotic forces acting which would cause the observed osmotic pressure to deviate from that pressure which should result from purely osmotic forces acting across the membrane. Depending upon the sign of the membrane potential and that of the ζ-potential, the observed osmotic pressure could be either higher or lower than that calculated on the basis of differences in concentration of the osmotically active components of the system on the two sides of the membrane. The difference between the observed osmotic pressure, P_o, and that calculated from the difference in osmotic concentrations, P_c, should then be definable in terms of the electrosmotic equation—

$$P_x = \frac{2E\zeta D}{\pi r^2},$$

where P_x is the electrosmotic pressure, which, if the theory holds, would be equal to $P_o - P_c$ at equilibrium; E is the potential across the membrane and is equal to the membrane potential derived from unequal distribution of ions across the membrane, being calculable, upon the basis of the Donnan theory, from the difference in active concentration of any one measurable ionic species (as H^+ ions); r is the radius of cross-section of the pore of the effective membrane; ζ is the electrokinetic potential at the solid-liquid interface within the pore of the effective membrane; and D is the dielectric constant of the liquid in the region of the interface.

A series of measurements of the osmotic pressure of gum arabic indicates the validity of this theory. The colloid used was gum arabic which had been purified and electrodialyzed free of all ions other than hydrogen ion. This electrodialyzed colloid acts as a strong acid (Thomas and Murray)[11] and each gram required 85×10^{-5} equivalents of alkali to neutralize it. Osmotic pressure measurements were obtained by placing known amounts of the colloid (the Na salt) in small collodion sacs and allowing equilibrium to be attained against a constant pressure applied to the solution inside of the sac while it was immersed in an external solution. The final concen-

tration of colloid within the sac was obtained by weighing the equilibrium contents of the sac. The total volume of the water in the system (inside and outside the sac) was known in each case and its distribution at equilibrium was obtained. Also the total equivalents of each diffusible ion in the system were known. After equilibrium had been attained, the H^+ ion concentration was determined electrometrically on the solutions from

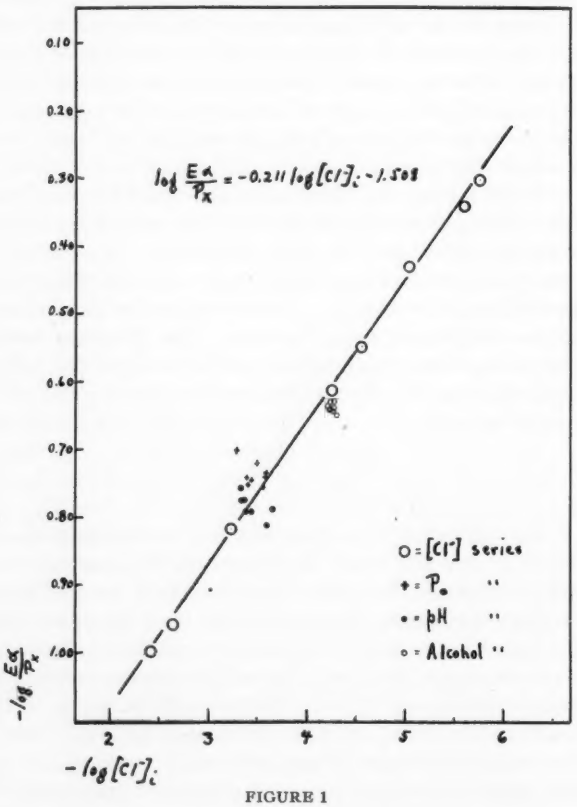


FIGURE 1

both sides of the membrane. From this determination the membrane potential, E, was obtained, as was also the ratio of distribution, R, of diffusible ions across the membrane. Then, on the basis of Donnan's equilibrium theory, it was possible to calculate the active concentrations of all diffusible ions on each side of the membrane. All experiments were made at 25°C., at which temperature one mole of dissolved material exerts an osmotic pressure of 25,280 cm. of H_2O. Neglecting any osmotic pressure

which might arise from the large colloid ion, the difference in the concentrations of diffusible ions multiplied by 25,280 gave the calculated osmotic pressure, P_c, which would result from the unequal distribution of these ions. The observed osmotic pressure, P_o, was equal to the constant applied pressure against which equilibrium had been attained.

The experiments included one series in which the observed pressure was varied and the external solution was common to all the sacs, one in which the pressure was constant and the pH was varied from 2 to 7 while the total equivalents of Cl^- ion in the system was constant, and a third series in which the pressure was constant and the concentration of NaCl in the system was varied from a very low value up to 0.01 N. In all experiments the system was kept on the acid side by addition of HCl in order to eliminate CO_2 absorption.

In these experiments the only diffusible ions in the system were H^+, Na^+ and Cl^-. In every case the value of P_c was greater than P_o. The values of E, as derived from pH measurements, where E (in millivolts) = 59.1 ($pH_o - pH_i$), were always such that the external solution was negative to the internal solution. Since the electrokinetic potential in the pore must be such that the wall was negatively charged to the liquid (whether the pore was that of the collodion or that of a gum arabic membrane the wall would be negatively charged since both are acidic and have no isoelectric point and NaCl or HCl cannot reverse the charge) it appeared that the electrosmotic pressure, P_x, should act counter to the true osmotic pressure P_c as the value of $P_c - P_o$ indicated it did.

In the electrosmotic equation given above, D could be taken as constant. In these and earlier experiments it was found that r could also be taken as constant. Whether the membrane sac was one of collodion, cellophane or parchment, made no difference in the final equilibrium concentration of colloid and H^+ ion distribution. The pore size of the supporting membrane could be varied widely with no effect upon these values. It appears that the effective membrane in all cases consisted of a layer of gum arabic laid down on the inner wall of the supporting membrane. For this reason the value of r effective in the electrosmotic equation was that of the gum arabic membrane and was therefore constant for all these experiments. That the effective membrane consisted of the gum arabic made possible estimates of the value of ζ, also, as seen below.

By collecting all constant terms of the equation on one side, the equation becomes

$$\frac{E\zeta}{P_x} = \frac{\pi r^2}{2D} = K.$$

Of the variables, E is taken equal to the membrane potential and P_x equal to $P_c - P_o$. An absolute value of ζ was not obtainable from the data

derived from these experiments, but by regarding the effective pores as those of the gum arabic membrane and assuming that ζ would be a function of α, the degree of ionization of the primary salt valences of the colloid, it was possible to obtain comparative values of ζ. It was assumed, also, that the value of α in the bulk of the colloid solution would equal the value of α for that part of the colloid making up the effective membrane. The equivalent weight and final concentration of the colloid were known. So were the final concentrations of H^+, Na^+ and Cl^+ in the colloid containing solution (from calculations described above). It was possible then to calculate the fraction of the colloid which was ionized, that is, α.

TABLE 1

P_o Cm. H_2O	P_c Cm. H_2O	P_x Cm. H_2O	$[H^+]_i$ $\times 10^7$	$[Na^+]_i$ $\times 10^7$	$[Cl^-]_i$ $\times 10^7$	E Mv.	α	$\frac{E\alpha}{P_x}$	$[Cl^-]_i^{0.211}$	$\frac{E\alpha[Cl^-]_i^{0.211}}{P_x}$
27.3	54.2	26.9	22950	24260	5000	28.7	0.1855	0.1980	0.2010	0.03980
37.8	70.7	32.9	26420	27930	4343	32.4	0.1844	0.1815	0.1950	0.03540
48.7	82.4	33.7	28830	30490	3980	34.7	0.1717	0.1767	0.1915	0.03380
85.7	117.8	32.1	36050	38110	3184	40.3	0.1522	0.1909	0.1828	0.03490
97.4	134.8	37.4	39430	41700	2914	42.6	0.1531	0.1744	0.1795	0.03130
110.5	147.5	37.0	41960	44350	2735	44.2	0.1519	0.1815	0.1770	0.03210
119.5	157.1	37.6	43830	46360	2617	45.3	0.1520	0.1832	0.1753	0.03210
127.1	157.5	30.4	53320	47320	4580	39.6	0.1348	0.1756	0.1972	0.03460
127.1	164.2	37.1	42170	61520	4505	40.1	0.1567	0.1694	0.1968	0.03330
127.6	171.2	43.6	30280	76400	4398	40.9	0.1805	0.1694	0.1958	0.03315
127.5	186.2	58.7	19630	93400	4191	42.0	0.2262	0.1620	0.1940	0.03140
128.1	201.0	72.9	9930	109470	4039	43.3	0.3025	0.1796	0.1922	0.03450
128.8	250.0	121.2	4062	136000	3558	46.5	0.4211	0.1618	0.1870	0.03025
129.5	388.3	258.8	933.0	194500	2503	55.9	0.7170	0.1550	0.1736	0.02690
129.0	472.0	343.0	16.1	229500	2200	59.2	0.9470	0.1637	0.1690	0.02765
131.5	320.5	189.0	49.5	129800	17.4	114.4	0.822	0.4975	0.0624	0.03105
130.8	344.1	213.3	62.2	139800	25.8	111.0	0.875	0.4561	0.0665	0.03031
131.5	351.5	220.0	74.1	146300	92	94.8	0.859	0.3701	0.0875	0.03240
131.2	381.9	250.7	171	165950	309	81.5	0.876	0.2850	0.1128	0.03215
131.0	398.5	267.5	235	177000	568	74.4	0.873	0.2430	0.1282	0.03114
130.4	278.4	148.0	1905	163200	5540	43.2	0.522	0.1523	0.2080	0.03168
130.4	217.0	86.6	7672	190400	23130	27.2	0.349	0.1097	0.2792	0.03062
130.1	195.2	65.1	12020	215300	39530	22.2	0.295	0.1005	0.3126	0.03141

If the theory holds that the pressure, $P_c - P_o = P_x$, could be defined by the electrosmotic forces acting, then the value $\frac{E\zeta}{P_x}$ would be a constant throughout all the experiments (disregarding any osmotic pressure derived from the colloid micellae). If ζ was a function only of α then also the value $\frac{E\alpha}{P_x}$ should be constant.

In the table of results, it is seen that, where the $[Cl^-]_i$ value was held nearly constant, the value $\frac{E\alpha}{P_x}$ was equal to a constant. When, however,

the foreign salt concentration in the colloid solution was changed widely the value $\frac{E\alpha}{P_x}$ varied in a manner such that at higher concentrations of salt (designated as $[Cl^-]_i$) $\frac{E\alpha}{P_x}$ became lower. That this effect was not due to a change in r, the radius of cross-section of the pore, as a result of dehydration of the micellae making up the effective membrane, was ruled out by adding a nonelectrolyte, ethyl alcohol, in place of NaCl. The concentration of alcohol was varied over a range 10 times as great as the range of concentration in which the effect of NaCl was observed and no variation in $\frac{E\alpha}{P_x}$ occurred. It became obvious that the observed effect on $\frac{E\alpha}{P_x}$ with increased salt concentration was an effect upon ζ. That is, ζ was a function of α but also a function of the foreign salt concentration in the solution. This latter function of ζ was presumably derived from oriented adsorption of the salt by the colloid and should therefore be definable in terms of the adsorption isotherm and ζ would be a function of the amount, a, adsorbed. The value of (a) could not be derived from the data at hand but the manner in which (a) varied with the salt concentration could be obtained by plotting $\frac{E\alpha}{P_x}$ against $[Cl^-]_i$ and a comparative value of (a) could be obtained in terms of $[Cl^-]_i$. In the graph shown, $\log \frac{E\alpha}{P_x}$ is plotted against $\log [Cl^-]$ and from the equation for the straight line thus obtained it appears that $\frac{E\alpha[Cl^-]_i^{0.211}}{P_x}$ is a constant. (This value for the earlier experiments is also plotted in the figure and shown in the last column of the table.)

It appears, then, that ζ is a function of α, the degree of ionization of the primary salt valencies of the colloid, and of (a) the degree of oriented adsorption of the foreign salt by the colloid. In this latter function of ζ appears the possibility of the lyotropic series playing a part in the electrosmotic forces acting, because oriented adsorption is a function of the relative adsorbability of both cation and anion of the salt and will not be the same for any two ions of equal sign and valence.

Without direct measurement of ζ but upon the assumption that ζ is a function only of α and (a), as the calculations indicate, it becomes apparent that anomalous osmosis observed for systems containing the colloid, gum arabic, can be fully defined in terms of the electrosmotic forces existing therein. The osmotic pressure derived from the colloid itself is so small as to be undetectable.

It is obvious that electrosmosis is an important osmotic factor which must be evaluated in every case where the osmotic properties of colloid solutions are being studied, and that its evaluation will furnish a satis-

factory explanation to most cases of anomalous osmosis in such systems.

Experiments with the Ca salt of arabic acid and with increasing concentration of $CaCl_2$ indicate that at a low concentration of the divalent cation salt the sign of ζ is reversed and negative osmosis occurs, i.e., the value of P_o is greater than P_c. This and lyotropic effects are being studied at the present time.

[1] Girard, P., *Compt. rend.*, **146,** 927 (1908); **159,** 99 (1914).
[2] Loeb, J., *J. Gen. Physiol.*, **1,** 717 (1919); **2,** 173, 255, 387, 563, 577, 659, 673 (1920).
[3] Bartell, F. E., *Colloid Symp.* Monograph, **1,** 120 (1923).
[4] Donnan, F. D., *Zeitschr. Electrochem.*, **17,** 572 (1911).
[5] Loeb, J., *J. Am. Chem. Soc.*, **44,** 1930 (1922).
[6] Bjerrum, N., *Z. f. phys. Chem.*, **110,** 656 (1924).
[7] Adair, G. S., *Proc. Roy. Soc.*, **A120,** 573 (1928).
[8] Hammarsten, E., *Biochem. Zeitschr.*, **144,** 383 (1924).
[9] Samac, M., Knop, L., and Pankovic, Z., *Koll. Zeitschr.*, **59,** 266 (1932).
[10] Freundlich, H., *Ibid.*, **18,** 11 (1916).
[11] Thomas, W. A., and Murray, H. A., *J. Phys. Chem.*, **32,** 676 (1928).

NOTE ON THE STURM-LIOUVILLE EIGENVALUE-EIGENFUNCTION PROBLEM WITH SINGULAR END-POINTS

BY EDWIN C. KEMBLE

RESEARCH LABORATORY OF PHYSICS, HARVARD UNIVERSITY

Communicated June 7, 1933

The general systematic theory of Sturm-Liouville (S.L.) eigenvalue-eigenfunction problems as discussed in standard texts such as Courant-Hilbert and Ince is restricted to cases involving regular end-points, finite fundamental intervals and homogeneous boundary conditions. The most important S.L. problems of mathematical physics lie outside this restricted class, however, and the question of the existence of discrete eigenvalues as well as that of the completeness of the system of corresponding eigenfunctions is usually dealt with for each problem separately. The general case of an S.L. problem with one regular boundary point in the finite region and a singular boundary point at infinity has been treated by Weyl[1] in a well-known paper whose primary objective was the development of an expansion theorem for the continuous spectrum. The class of problems directly treated in Weyl's paper is not perfectly general and his technique is too elaborate to be readily accessible to physicists. Hence the writer has considered it worth while to develop a general treatment of the discrete spectra of S.L. problems with singular end-points which involves only elementary methods. The procedure is only sketched in this note which the author hopes to follow with a more detailed publication elsewhere.

Consider the S.L. equation

$$\frac{d}{dx}\left[p(x) \frac{dy}{dx}\right] - q(x)y + \lambda\rho(x)y = 0 \qquad (1)$$

in which λ is a parameter while $p(x)$, $q(x)$, $\rho(x)$ are real continuous functions of x in the fundamental interval $a < x < b$. $p(x)$ and $\rho(x)$ are required to be positive at every interior point of this interval. The boundary points a, b are singular and either or both of these points may lie at infinity. The solution of this equation becomes an eigenvalue-eigenfunction problem if we impose suitable boundary conditions. Those defined by physical applications vary from case to case and it is therefore convenient to introduce a set of standard mathematical boundary conditions to which the various physical conditions can usually be reduced.

To this end we require that there shall exist a function $g(x)$ such that

$$\operatorname*{Lim}_{x \longrightarrow a} \left\{ p(x) \left| y(x) \right|^2 g^{-1}(x) \right\} = 0; \qquad (2)$$

$$p(x) \left| y'(x) \right|^2 g(x) < M; \quad 0 < x - a < B \qquad (3)$$

where M and B are positive quantities independent of x. $g(x)$ is to be independent of the parameter λ whose eigenvalues we seek, and can usually be set equal to unity. This condition we call the *singular point boundary condition* (s.p.b.c.) for the end-point a. A corresponding condition is to be applied to b.

The s.p.b.c. has the following important properties.

(a) If $u_1(x)$ and $u_2(x)$ conform to the s.p.b.c. at $x = a$

$$\operatorname*{Lim}_{x \longrightarrow a} \left\{ p(x) u_1 u_2^{*\prime} \right\} = \operatorname*{Lim}_{x \longrightarrow a} \left\{ p(x) u_1' u_2^{*} \right\} = 0. \qquad (4)$$

(The asterisk denotes the complex conjugate.)

(b) If $u_1(x)$ and $u_2(x)$ conform to the s.p.b.c. at $x = a$, any linear combination, say, $w(x) \equiv \alpha u_1 + \beta u_2$, will conform to the same condition. This is a consequence of the inequality

$$\left| w \right|^2 < 2 \left\{ \left| \alpha u_1 \right|^2 + \left| \beta u_2 \right|^2 \right\}.$$

(c) Let $y_1(x)$ and $y_2(x)$ denote two linearly independent integral curves of Eq. (1) constituting a fundamental system. Then both of these functions cannot conform to the s.p.b.c. at either end-point, for it follows from the linear independence that the Wronskian

$$W(y_1, y_2) \equiv p(y_1 y_2' - y_1' y_2) \neq 0,$$

whereas if both y_1 and y_2 satisfy the s.p.b.c. at either end-point, it follows from (a) that W must vanish at the corresponding boundary point.

An important corollary on (a) is that any two eigenfunctions $y_1(x)$,

$y_2(x)$ with different eigenvalues λ_1 and λ_2 must satisfy the orthogonality relation

$$\int_a^b \rho y_1 y_2^* dx = 0.$$

From (*c*) we draw the further inference that there cannot be two linearly independent eigenfunctions with the same eigenvalue.

We approach the question of the existence of eigenvalues by postulating the existence of an interval G of λ-values, say, $\lambda' \leqslant \lambda \leqslant \lambda''$ such that for every λ in G there exists a pair of integral curves $u_\lambda(x)$, $v_\lambda(x)$ which conform to the s.p.b.c. at a and b, respectively, and which do not have an infinite number of nodes in the neighborhood of these points. It follows from this assumption that there exist continuous functions of x and λ, say, $u(x,\lambda)$, $v(x,\lambda)$ which for any given λ in G satisfy the conditions laid down for u_λ and v_λ.

If the Wronskian $W[u(x,\lambda), v(x,\lambda)]$ vanishes for any value of λ in G, say, λ_1, $u(x,\lambda_1)$ and $v(x,\lambda_1)$ are linearly dependent and conform to the s.p.b.c. at both boundary points. Hence λ_1 is an eigenvalue and $u(x,\lambda_1)$, $v(x,\lambda_1)$ are eigenfunctions. It is easy to prove by standard methods that as λ increases the nodes of $u(x,\lambda)$ move continuously to the left while those of $v(x,\lambda)$ move to the right. Whenever a node of u comes into coincidence with a node of v, they all coincide in pairs, the Wronskian vanishes and we have an eigenvalue.

Let n'' be the number of interior nodes of $u(x,\lambda'')$ [upper limit of G] and let n' be the number of interior nodes of $u(x,\lambda')$. For simplicity we assume that λ'' and λ' are not eigenvalues. It follows readily from the above argument that there exist $n'' - n'$ eigenvalues in the interval G, each having an eigenfunction with one more interior node than the eigenfunction of next lower eigenvalue. The case where $n' = 0$ requires special discussion but forms no exception to the rule. If $n'' > 0$, and if the lower limit of G is at $\lambda = -\infty$, there is a minimum eigenvalue with a nodeless eigenfunction. If the upper limit of G is at $\lambda = +\infty$, the sequence of ascending eigenvalues is denumerably infinite and $\operatorname{Lim}_{n \to \infty} \lambda_n = \infty$.

The S.L. eigenvalue-eigenfunction problem is usually reducible to variational form. One of these forms consists in finding stationary values of quadratic integral

$$J[y,\lambda,a,b] = \int_a^b \{p|y'|^2 + q|y|^2 - \lambda\rho|y|^2\}dx, \qquad (5)$$

the admissible comparison functions being subject to boundary conditions which insure that

$$\operatorname{Lim}_{x \to a} \{p[\delta y y^{*\prime} + \delta y^* y']\} = \operatorname{Lim}_{x \to b} \{p[\delta y y^{*\prime} + \delta y^* y']\} \qquad (6)$$

and to the requirement that they yield convergent values of J. The Euler's equation of (5) is (1), the s.p.b.c. take care of (6) and the eigenfunctions of (1) yield the value zero for J. If they also yield convergent values of the normalization integral $N = \int_a^b \rho \mid y \mid^2 dx$, they must yield convergent values of $Q = \int_a^b (p \mid y' \mid^2 + q \mid y \mid^2) dx$. In this case the eigenvalues of λ are stationary values of Q/N for comparison functions subject to the same restrictions as before plus the requirement that N shall converge. In applications to wave mechanics this latter requirement is usually one of the physical conditions to which the s.p.b.c. are assumed equivalent.

Changing the dependent variable from y to w by the transformation $w = p^{1/2}y$ converts the variational problem into one involving fixed end-points, since $w(a) = w(b) = 0$. It is a theorem of the calculus of variations that if $y = U(x)$ is a nodeless extremal of an integral $D[y] = \int_{x_0}^{x_1} F(x,y,y')dx$ of the form (5) subject to fixed end-point conditions, and if $F_{y'y'}(x,y,\beta) > O$ for every point (x,y) in the neighborhood of $U(x)$ and for every finite β, then U actually minimizes D.[2] In proving this theorem the end-points x_0 and x_1 are assumed regular. Were it not for this circumstance we could identify $J[y,\lambda,a,b]$ with D and show that when Eq. (1) has a nodeless extremal y_0 yielding a convergent $N[y_0]$ the corresponding eigenvalue λ_0 is a minimum for Q/N.

To overcome the difficulty regarding the singularities at the end-points it is necessary to introduce an infinite sequence of modified variational problems $A_1, A_2, \ldots A_n, \ldots$ with regular end-points a_n, b_n inside the fundamental interval $a < x < b$ and with fixed end-point conditions. This sequence can be so chosen that the lowest eigenvalue and eigenfunction converge on λ_0 and $w_0 = p^{1/2}y_0$ as n becomes infinite. By application of the above theorem to the members of this sequence we can prove that y_0 actually minimizes Q/N. Moreover, the minimum is absolute since for quadratic integrals of the type under consideration the increment due to any variation in the argument function is equal to the sum of the first and second variations of the integral and hence is positive or zero.

The higher eigenvalues of Eq. (1) can be treated in the same way,[3] each being proved to yield the minimum value of Q/N for all comparison functions which obey the s.p.b.c., yield convergent integrals and are orthogonal to all eigenfunctions of lower eigenvalue. When this point is established, one can adopt the procedure of Courant-Hilbert[4] to show that when the interval G extends from λ_0 (or a lower value of λ) to ∞, the eigenfunctions form a *complete set.*

The question of the range of the discrete spectrum of eigenvalues of Eq. (1) and the completeness of its system of eigenfunctions is thus reduced

to the question of the validity of our postulates regarding the equivalence of the physical boundary conditions to the s.p.b.c. and the existence and range of the interval G. These points must be investigated separately for each individual case, using primarily the standard methods for the study of the behavior of solutions of a differential equation in the neighborhood of a singular point.

The writer is greatly indebted to his colleague, Prof. Marston Morse, for numerous suggestions and much helpful discussion in connection with the proof that the eigenvalues of λ are minima of Q/N for suitably restricted classes of admissible comparison functions.

[1] Weyl, Hermann, *Math. Ann.*, **68**, 220 (1910).

[2] Bolza, O., *Lectures on the Calculus of Variations*, Chicago, p. 96 (1904).

[3] Kuen Sen Hu, Thesis, Theorem 10.3 [*Contributions to the Calculus of Variations*, Univ. of Chicago Press, 1933].

[4] Courant, R., and Hilbert, D., *Methoden der Math. Physik*, Berlin, 1931, Kap. 6, § 3.

STUDIES ON THE GROWTH HORMONE OF PLANTS. III. THE INHIBITING ACTION OF THE GROWTH SUBSTANCE ON BUD DEVELOPMENT

By Kenneth V. Thimann and Folke Skoog

William G. Kerckhoff Laboratories of the Biological Sciences, California Institute of Technology, Pasadena, California

Communicated April 24, 1933

It has long been known that when the growing point of a young dicotyledonous plant is removed, the axillary buds on the stem below it begin to develop. As long as the terminal bud is present, the development of the axillary buds is inhibited. A lateral bud may also be inhibited by the rapid growth of another lateral above it, as we have found in our experiments, or opposite it, as in those of Dostál (1926). Furthermore, Snow (1929a) has shown that the inhibition, in *Pisum*, is principally due to the young leaves in the developing bud. The evidence indicates that this inhibition is probably caused by a special substance (see Snow, 1929b). There was reason to believe that this inhibiting substance is of the same nature as the growth-promoting substance of *Avena* coleoptiles, and the experiments to be described here confirm this belief.

1. *Methods.*—Young plants of *Vicia faba*, 4 to 6 weeks old, grown in the light, were used. The lowest lateral buds of these plants grow with great regularity. The plants selected were of approximately equal height and with as nearly as possible equal numbers of buds and leaves. The

growth substance used was obtained from the growth of *Rhizopus suinus* (Thimann and Dolk, 1933) and had an activity of about 2.10^{-6} mg. per plant unit. The activity of this growth substance and of that obtained from the *Vicia* plants was tested on *Avena* coleoptiles under the standard conditions described by Went (1928), using the definition of units as given by Dolk and Thimann (1932).

2. *Production of Growth Substance by Vicia faba.*—It was first found that the terminal bud produces growth substance in rather large quantities. From terminal buds of young plants 12 cm. high, 30 to 40 plant units diffused out into agar blocks in an hour. From the buds of older plants the amount obtainable was less. The undeveloped lateral buds produce practically no growth substance. When, however, the plant is decapitated and the lateral buds are undergoing rapid development, each bud produces almost one-half as much growth substance as the terminal bud of an intact plant. Small amounts of growth substance are also produced by the leaves, less than one plant unit per hour being obtainable from the oldest leaves, and about 5 plant units per hour from the youngest leaves under the conditions of the experiment. It is therefore clear that growth substance production is associated with the actively developing parts of the plants, and that rather large quantities of the substance are regularly passing into the stem.

3. *Inhibition of Bud Development by Growth Substance.*—In order to make quantitative determinations of the effect of applying growth substance to the plant, it was necessary first to determine the time required for growth substance to enter the plant from an agar block. By plotting the amount of growth substance remaining in agar blocks after different periods of application to the stem, it was found that 6 hours were necessary for a complete transference of growth substance into the stem.

Agar blocks containing growth substance were therefore applied to the tops of decapitated plants every 6 hours. Intact plants, and also plants to which blocks of plain agar were being applied, were used as controls. Measurements of the length of the lateral buds, and also of the main stem, were made daily. When the amount of growth substance applied was of the same order as that diffusing from the terminal bud into agar (160 plant units every 6 hours), a slight but definite inhibition of the development of the lateral buds was observed. When, however, the amount of growth substance applied was larger than that diffusing from the terminal bud (1400–1700 plant units every 6 hours), the development of lateral buds was completely inhibited. Thus, in one experiment, the mean increase in length of the lateral buds in 8 days was 3.4 ± 0.5 mm. in intact plants, 16.0 ± 2.7 mm. in decapitated controls and 1.8 ± 0.6 mm. in plants to which 1670 units of growth substance had been applied. Several such experiments were carried out, each comprising thirty to forty

buds, and the amount of inhibition was similar in each case. The inhibition also takes place in the dark.

The necessity for applying larger amounts of growth substance than can be obtained from the terminal bud is fully justified on the ground that the application is generalized over the whole stem surface, while the normal supply from the tip is localized in the conducting tissue and therefore more effective in its action. Furthermore, there is evidence that wound substances may inactivate a part of the applied growth substance (Kisser, *et al.*, 1931). The inhibition is not to be ascribed to any damage, since the plants remained in good condition throughout the experiment, as was shown by the rapid development of the lateral buds as soon as the application of growth substance was stopped. While it is possible that the inhibition is due, not to the growth substance, but to another compound of similar nature present in the active concentrates, the present experiments make it probable that it is the growth substance itself, produced in the terminal bud, which inhibits lateral bud development.

Although it seems paradoxical that a substance promoting cell extension can also act as an inhibitor, this fact provides an explanation for much of the earlier work on inhibition, such as the experiments of Dostál, Snow (1929a) and Weiskopf (1927). The probable mechanism of the effect, together with a more detailed account of these experiments, will be published elsewhere.

Dolk, H. E., and Thimann, K. V., *Proc. Nat. Acad. Sci.*, **18**, 30 (1932).
Dostál, R., *Acta. Sci. Nat. Morav.*, **3**, 83 (1926).
Kisser, J., Stasser, Kiffe and Göllner, *Anz. Akad. Wiss. Wien*, 275 (1931).
Snow, R., (*a*) *New Phytologist*, **28**, 345 (1929); (*b*) *Annals of Botany*, **43**, 261 (1929).
Thimann, K. V., and Dolk, H. E., *Biol. Zentr.*, **53**, 49 (1933).
Weiskopf, B., *Publ. biol. l'école hautes études vétérinaires Brno*, **6**, No. 4 (1927).
Went, F. W., *Rec. trav. bot. Néerl.*, **25**, 1 (1928).

STUDIES ON THE GROWTH HORMONE OF PLANTS. IV. ON THE MECHANISM OF THE ACTION

BY JAMES BONNER

WM. G. KERCKHOFF LABORATORIES OF THE BIOLOGICAL SCIENCES, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated May 19, 1933

Introduction.—It has been known for some time that among the higher plants cell elongation as well as tropistic response is governed by a well-defined chemical substance which is secreted by young leaves and by the apical portions of stems.[1,2] The mechanism by which this substance exerts its effect upon plant cells has, however, remained obscure and it is relatively recently that an attack upon this problem has been made.

In an earlier paper it was shown that the amounts of this "growth substance" which enter the *Avena* coleoptile are far too small to bear any simple stoichiometrical relation to the substances which are formed in the cell wall during the resultant growth.[3] The growth substance is therefore a true hormone in that it is active in minute amounts and in some indirect manner. The work of Heyn,[4] of Heyn and van Overbeek,[5] and of Söding[6] has shown that *Avena* coleoptiles which are artificially supplied with growth substance exhibit a marked increase in the plasticity of their cell walls. From this result it seemed possible that the growth substance itself might possess the property of being able, merely by its presence, to decrease the viscosity of the fluid in which the cellulose micelles of the cell wall are embedded. The cell wall could then be readily stretched by the force at hand, which would, under normal conditions, be the force due to the osmotic pressure of the cell contents. It was found in this laboratory, however, that if the osmotic force be replaced by the force due to a suitable weight, no increase in plasticity of the coleoptile results from the application of growth substance. It seemed, therefore, more probable that the hormone exerts its influence upon the cell wall through the mediation of the cell protoplasm. The present paper will describe briefly experiments which demonstrate that this is actually the case. A fuller account will appear elsewhere.

Materials and Methods.—*Avena* plants of the pure line, "Sieges Hafer," kindly supplied by Dr. Åkermann of Svalöf, were used throughout. The plants were grown under the usual rigidly controlled conditions of temperature and humidity.[9] All experiments were carried out under red light, to avoid photropic stimulation of the plants.

The conventional method of applying growth substance dissolved in agar blocks to the stumps of decapitated *Avena* coleoptiles is not well adapted to the study of matters concerned with metabolism. It was

found, however, that short sections (three mm. long) cut from the coleoptile and suspended in growth substance of suitable concentration elongated rapidly, even in the absence of their usual food source (i.e., the seed). The various substances whose effects upon the action of growth substance are to be investigated may be conveniently added to the growth substance solution in which the sections are suspended. In order to obtain sections which were initially relatively free of growth substance, the hormone-producing tip was removed from each plant two hours before the short sections were cut from the coleoptile.

Experimental.—Avena coleoptiles have not previously been grown, as in the present case, suspended in solutions of growth substance. It is therefore of interest to enumerate, as follows, some of the factors which were found to control elongation under these conditions:

(*a*) Sections immersed in water or solution of growth substance do not exhibit "regeneration" of the capacity to produce growth substance which is characteristic of decapitated plants grown in the usual manner with only their roots in water.

(*b*) Sections from near the apex of a previously decapitated coleoptile elongate more rapidly in growth substance of a given concentration than sections from near the base of the same plant. This indicates that cells which have already undergone considerable elongation are less reactive to the hormone than those which have elongated only slightly.

(*c*) The rate and extent of the elongation vary greatly with the growth substance concentration. There is an optimum concentration above which the hormone is toxic and may even bring about a loss of turgor.

The growth of coleoptile sections under the influence of growth substance was found to be completely inhibited by 10^{-3} normal KCN, and by 0.05 per cent phenyl-urethane. This inhibition suggested a relation between the action of growth substance and respiratory activity. Coleoptile sections were, therefore, immersed in a solution of growth substance under an atmosphere of nitrogen. The sections were not harmed by a prolonged lack of oxygen since they grew normally if placed in air. Nevertheless, neither elongation nor any action of growth substance preliminary to elongation took place in nitrogen. This result might be due either to a failure of the hormone to penetrate into the cells in the absence of oxidative processes, or to the participation of growth substance itself in an oxidative process. In the first case it would hardly be expected that the respiration of a section should be effected by the addition of the hormone, since Steward has shown that, in the case of potato tissue at least, the uptake of ions, while dependent upon aerobic metabolism, nevertheless does not increase the production of carbon dioxide.[7] In the second case it would be difficult to predict *a priori* whether or not an increase in respiratory activity would occur upon the addition of growth substance.

A preliminary study of the effect of growth substance upon the respiration of coleoptile sections was made, using the standard Warburg technique.[8] The results may be summarized as follows:

(*a*) Growth substance in low concentrations stimulates the respiration of coleoptile sections by as much as 27 per cent during the first two hours after its addition. In high concentrations growth substance greatly inhibits the respiration. There is apparently some discrepancy between the effects of hormone concentration on growth and on respiration, although the optimal concentrations are similar.

(*b*) The respiration of sections is inhibited by the same concentrations of KCN and of phenyl-urethane which were found to inhibit elongation under the influence of growth substance.

(*c*) The additional respiration observed upon the addition of growth substance exhibits the same sensitivity to KCN and to phenyl-urethane as does the normal respiration and is therefore probably identical with it.

(*d*) Growth substance which has been inactivated for growth by oxidation with hydrogen peroxide does not stimulate respiration.

Discussion.—From the foregoing results it seems possible that an increase in respiration is one member in the chain of processes by which growth substance brings about elongation. It is as yet unsafe to conclude that this is actually the case. In the first place, the hormone preparation used was, while of relatively high purity, still only approximately one per cent hormone. It is possible that an impurity in the preparation is responsible for the stimulation of respiration. In the second place, this stimulation may be a secondary phenomenon attending the presence of growth substance in the cell, and have in reality no bearing upon growth. These questions, as well as a possible relation of respiratory rate to cell wall plasticity, are now being investigated.

The author wishes to express his appreciation to Dr. Robert Emerson for his many suggestions and for the use of his laboratory, in which the measurements of respiration were made, and to thank Dr. F. Went and Dr. Kenneth V. Thimann for their help and suggestions.

[1] Went, F., *Rec. Trav. bot. neerl.*, **25**, 1 (1928).
[2] Kögl, F., Haagen-Smit, A., and Erxleben, H., *Zeit. Physiol. Chem.*, **214**, 241 (1933).
[3] Thimann, K. V., and Bonner, J., *Proc. Roy. Soc. London.* (In press.)
[4] Heyn, A., *Rec. Trav. bot. neerl.*, **28**, 113 (1931).
[5] Heyn, A., and van Overbeek, J., *Proc. kon. Akad. Wetensch. Amster.*, **34**, 1190 (1931).
[6] Söding, H., *Jb. Wiss. Bot.*, **74**, 127 (1931).
[7] Steward, F., *Protoplasma*, **18**, 208 (1933).
[8] Warburg, O., *Stoffwechsel der Tumoren*, Berlin, (1926).
[9] Dolk, H., and Thimann, K. V., these PROCEEDINGS, **18**, 30 (1932).

THE ENERGY OF UREA SYNTHESIS. II. THE EFFECT OF VARYING HYDROGEN ION CONCENTRATION WITH DIFFERENT METABOLITES

BY HENRY BORSOOK AND GEOFFREY KEIGHLEY

WILLIAM G. KERCKHOFF LABORATORIES OF THE BIOLOGICAL SCIENCES, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated June 13, 1933

In a previous communication[1] we have reported that an increased oxygen consumption can be observed accompanying the synthesis of urea by liver slices from ammonium bicarbonate. This observation is now confirmed, and the results of additional experiments are presented which elucidate to some extent some of the general features of the mechanism of the transfer of energy in this and probably other coupled reactions.

The technique employed here was similar to that in the experiments reported previously, except that the vessels used were two and one-half times smaller, and the liver slices were placed in 3 cc. instead of 15 cc. of fluid. The composition of the Ringers solution was varied according to the pH desired. The hydrogen ion concentrations were measured colorimetrically before and after each experiment. The ornithine and the ammonium chloride, wherever it was used, were first brought to pH 7.4 with sodium hydroxide. Q_{O_2} is the number of cubic millimeters of O_2 at standard temperature and pressure used per milligram of dry weight of tissue (dried at 100°C.) per hour. Q_{urea} is the number of cubic millimeters of CO_2, at standard temperature and pressure, liberated at the end of the experiment at pH 5.0 by urease per milligram of dry weight of tissue per hour.

Some of the results obtained are given in the table. They show that with no metabolite added or with glucose, at pH 7.4 and 8.4 one molecule of additional oxygen is used for every molecule of urea synthesized. With dl lactate and with succinate at pH 8.4 the results were the same. On the other hand, at pH 7.4 with the latter two metabolites there was much less additional oxygen used in the synthesis of urea. These observations resolve the apparent discrepancy between our previous observations and those of Krebs and Henseleit.[2]

The latter authors worked for the most part at pH 7.4. The one experiment whose protocoll on oxygen consumption is given was carried out at this pH with dl lactate as metabolite. The Ringers solution was the same as in experiments 1 and 2 above, and a gas mixture containing 5 per cent CO_2 and 95 per cent O_2 was used. The CO_2 was not absorbed and so maintained the hydrogen ion concentration of the Ringers solution at pH 7.4. In our experiments the gas was pure O_2, and the CO_2 was continually ab-

sorbed. As a result, in our previous experiments, a large fraction of the bicarbonate was decomposed, producing an alkaline reaction until finally a balance was attained between the CO_2 lost by the solution and that produced by the respiration of the tissue. We were, of course, aware of this induced alkalinity, but these conditions were preferred in those exploratory experiments because they permitted us to observe the rate of respiration throughout the experiment; and because larger amounts of urea are synthesized under these conditions than when the pH was maintained at 7.4 by a phosphate buffer. Warburg has shown that between pH 6.5 and 8.5 the respiration of liver slices is not affected by change in hydrogen ion concentration,[3] which our own experience confirmed.

In order to obviate a lengthy discussion of the different hypotheses which might be invoked to explain these results, we shall describe briefly the results of some experiments in a different field which bear directly upon the question of the mechanism of the transfer of energy in biological coupled reactions.* In these experiments it was found that in order to obtain under anaerobic conditions the synthesis of lactate from pyruvate through the oxidation of formate by toluene treated bacteria, it was necessary to add a small amount of some substance (usually a dye) which is capable of oxidation or reduction by both the formic acid and lactic acid dehydrogenase mechanisms. In the absence of the dye the formic acid was not oxidized and the pyruvic acid was not reduced. Similar results were obtained with the formation of succinic acid from fumaric acid at the expense of the oxidation of lactic acid to pyruvic acid. The transfer of energy here was effected by the dye, which passed in the reduced state from the locus where, for example, formic acid was oxidized and the dye reduced, to the center where, by oxidation of the dye, pyruvic acid was reduced to lactic acid.

The simplest hypothesis accounting for all of the facts at present seems to us as follows: The coupled reaction in which urea is synthesized from ammonia is essentially the same with all the metabolites used, at pH 7.4, as well as at pH 8.4; because in all cases the synthesis of urea is accelerated by ornithine; and because with lactate and succinate at pH 8.4 the ratio of increased oxygen used to urea synthesized is the same as with glucose or no added metabolite at pH 7.4 and pH 8.4. Here the energy produced is nearly 10 times the amount necessary, and is derived from specific oxidations and not from the general respiration of the cell. The synthesis, i.e., the transfer of energy occurs only at certain specific centers. This is shown by the absence of any correlation between the total rate of metabolism, in the absence of urea synthesis, and the amount of urea synthesized when ammonia and ornithine are added. This is further sup-

* These experiments were carried out with Mr. H. F. Schott in this laboratory, and will be submitted for publication shortly.

RELATION BETWEEN ADDITIONAL OXYGEN CONSUMPTION AND UREA SYNTHESIS BY LIVER TISSUE IN RINGERS SOLUTIONS AT DIFFERENT HYDROGEN ION CONCENTRATIONS AND WITH DIFFERENT METABOLITES

EXPERIMENT NO.	SOURCE OF AMMONIA	TYPE OF RINGERS SOLUTION USED	pH	ORNITHINE CONCENTRATION MG. %	AMMONIA CONCENTRATION MG. %	DRY WT. OF TISSUE MG.	DURATION OF EXPERIMENT MINUTES	O_2 CONSUMPTION Q_{O_2}	UREA FOUND Qurea	ΔQ_{O_2}	ΔQurea	$\frac{\Delta Q_{O_2}}{\Delta Q\text{urea}}$
				10	..	13.4		9.0	0.3			
7	NH_4Cl	0.2% Glucose in phosphate Ringers	7.3				87			2.6	2.7	1.0
				10	25	12.4		11.6	3.0			
				10	25	20.6		10.7	2.1			
8	NH_4Cl	0.2% Glucose in phosphate Ringers	7.4				87			2.3	2.0	1.2
				10	..	20.2		8.4	0.1			
				10	25	12.9		16.2	6.2			
1	NH_4HCO_3	Bicarbonate Ringers	8.4				85			5.7	5.9	1.0
				10	..	13.0		10.5	0.3			
				10	25	13.7		16.0	5.8			
2	NH_4HCO_3	Bicarbonate Ringers	8.4				85			5.6	5.3	1.05
				10	..	14.3		10.4	0.5			
				10	..	15.8		11.3	0.3			
21	NH_4Cl	Phosphate Ringers	7.4				85			1.4	1.6	0.9
				10	25	13.9		12.7	1.9			
				10	..	13.7		10.2	0.2			
22	NH_4Cl	Phosphate Ringers	7.4				85			1.6	1.6	1.0
				10	25	13.6		11.8	1.8			
				10	25	14.5		16.6	5.2			
19	NH_4Cl	Bicarbonate Ringers	8.2				91			4.1	4.7	0.9
				10	..	14.4		12.5	0.5			

				10	25	14.9		16.5	5.3				
20	NH_4Cl	Bicarbonate Ringers	8.2				91			3.6	4.6	0.8	
				10	..	13.2		12.9	0.7				
				10	25	11.9		16.1	3.3				
10	NH_4Cl	0.2% dl lactate in phosphate Ringers	7.4				95			1.5	2.8	0.5	
				10	..	11.7		14.6	0.5				
				10	25	9.9		17.1	3.8				
11	NH_4Cl	0.2% dl lactate in phosphate Ringers	7.4							1.1	3.3	0.3	
						10.1	95	16.0	0.6				
				10	..	10.3		15.0	0.0				
12	NH_4Cl	0.2% dl lactate in bicarbonate Ringers	8.4				89			2.5	2.5	1.0	
				10	25	8.0		17.5	2.5				
				10	..	10.7		14.6	0.0				
13	NH_4Cl	0.2% dl lactate in bicarbonate Ringers	8.4				89			2.1	2.2	1.0	
				10	25	10.4		16.7	2.2				
				10	25	15.8		16.0	2.1				
23	NH_4Cl	$^1/_{90}$ molal succinate in phosphate Ringers	7.4				98			−0.5	1.9	0.0	
				10	..	14.2		16.5	0.3				
				10	25	14.3		16.9	2.1				
24	NH_4Cl	$^1/_{90}$ molal succinate in phosphate Ringers	7.4				98			−0.1	1.9	0.0	
				10	..	14.4		17.0	0.3				
				10	25	15.7		20.3	3.2				
17	NH_4Cl	$^1/_{90}$ molal succinate in bicarbonate Ringers	8.3				78			2.7	2.7	1.0	
				10	..	16.0		17.6	0.5				
				10	25	14.7		21.5	3.6				
18	NH_4Cl	$^1/_{90}$ molal succinate in bicarbonate Ringers	8.3				78			3.0	3.1	1.0	
				10	..	14.2		18.5	0.5				

ported by the experiments on the reductive formation of succinic acid and lactic acid mentioned above; and by the demonstration by Krebs and Henseleit that urea synthesis is obtained only with intact cells; when the structure of the liver cells is destroyed by grinding, though respiration still occurs there is no longer any synthesis of urea. This evidence for the existence of specific centers also strengthens the first point that the essential mechanism of the coupled reaction is the same under all conditions.

It follows therefore that with lactate and succinate at pH 7.4 (but not at pH 8.4) other reactions requiring energy are in progress. In these the energy supplied to them is greatly in excess of that required. When ornithine and ammonia are added some of this surplus energy, which is possibly used in other reactions or which passes off as heat, is used in the synthesis of urea. The result is that under these conditions the increased oxygen consumption accompanying the synthesis of urea is less than under other conditions, e.g., at pH 8.4 or with glucose instead of lactate or succinate, when the increased oxygen consumption is proportional to the amount of urea synthesized.

The synthesis of glycogen suggests itself as an obvious possibility for this first-mentioned reaction requiring energy upon which the synthesis of urea is superimposed. It has been shown that both lactate and succinate may be converted to glycogen in the liver.[4,5] In the case of the synthesis of glycogen from lactate it is well known that more energy is produced than the theoretical minimum necessary. It is probable that a similar state of affairs exists in the synthesis of glycogen from succinate.

It is necessary to the hypothesis advanced above that the metabolite added to the Ringers solution, particularly the lactate and succinate, be used. The results show that, in the absence of ammonia, at all the hydrogen ion concentrations tried, the rate of respiration with the latter two metabolites was much higher than when no metabolite, or when glucose was added.

It is clear from the above data that the total energy change is not governed by the energy requirements of the reaction in which ΔF is positive. The total quantity of energy produced is much greater than the necessary minimum amount computed from solely thermodynamical considerations. It is probable that this is a very common, if not a universal feature of biological coupled reactions. The production of a large amount of excess energy in the synthesis of glycogen from lactic acid may be cited as another example. The 1:1 ratio between additional oxygen used and urea synthesized, suggests that the factor governing the total energy change here is a stoichiometrical one, the consumption of a molecule of oxygen really representing the production of a molecule of CO_2 at a locus in the cell, and in such a state that it can be combined with ammonia and the other components necessary for the formation of urea.

Summary.—Experimental results are presented which indicate (1) that one molecule of oxygen is used for every molecule of urea synthesized; (2) that the total energy change is governed by the stoichiometrical rather than by the energy requirements of the reaction in which the free energy change (ΔF) is positive; (3) the existence of specific centers and specific mechanisms in the cell whereby energy is transferred from specific energy liberating reactions (i.e., not the general respiration of the cell) to the energy absorbing reaction; and (4) that this specificity does not exclude mechanisms whereby the energy produced by an energy liberating reaction may be shared by more than one energy absorbing reaction.

[1] Borsook, H., and Keighley, G. L., *Proc. Nat. Acad. Sci.*, **19**, 626–631 (1933).

[2] Krebs, H. A., and Henseleit, K., *Zeit. physiol. Chem.*, **210**, 33 (1932).

[3] Warburg, O., *Biochem. Z.*, **143**, 317 (1923).

[4] Himwich, H. E., Koskoff, Y. D., and Nahum, L. H., *J. Biol. Chem.*, **85**, 571 (1930); Bayo, C. Pi-Suñer, and Pi, J. F., *Biochem. Z.*, **242**, 306 (1931).

[5] Ponsford, A. P., and Smedley-Maclean, I., *Biochem. J.*, **26**, 1340 (1932).

---

*ON THE TEMPERATURE-REGULATORY FUNCTION OF "SPONTANEOUS" ACTIVITY IN THE MOUSE**

BY T. J. B. STIER

LABORATORY OF GENERAL PHYSIOLOGY, HARVARD UNIVERSITY

Communicated June 9, 1933

When every effort is made to eliminate excitation in the external environment of a young mouse, movements still occur—the so-called "spontaneous" movements. These movements do not occur haphazardly, but with a definite form, namely, in regularly recurring periods of movement alternating with periods of quiet. The periods of activity and of quiescence are of statistically constant length, having a mean probable error, as percentage of the mean, of 5 per cent for activity, of 10 per cent for quiescence, at 21 ± 0.01°C. (Stier, 1930). The mice are still poikilothermic at the age of two days; the frequency of occurrence of activity cycles increases with temperature in the manner of a chemical reaction velocity (μ, in the Arrhenius equation $\ln K \propto \mu/RT$, was found to be about 25,300 calories; cf. Stier, 1930). The duration of a period of activity, however, decreases as the body temperature increases, e.g., at 16.5°C. the average duration of a single activity period was 0.7 min., at 29.5°C. the average duration was 0.3 min.

The average duration of activity periods times the number of periods per hour yields a quantity which should be roughly proportional to the rate at which heat is produced by muscular movements of the young mouse.

Calculating this product for various body temperatures in the range 15° to 32.5°C. gives an indication of the heat contributed by the active muscles of the young mouse over the range of body temperatures tested (cf. Fig. 1). The internal temperatures were measured by a thermopile (cf. Stier, 1930; and Stier and Pincus, 1927–28). At the moment we do not know whether the rate of heat production at any one temperature remained constant during each period of activity. The kymograms of the spontaneous movements show a uniform vigor of movements during each part of a period of activity (Stier, 1930). We have also been unable, as yet, to find out how the "intensity" of muscular activity exhibited during single activity periods is related to temperature. Probably the frequency of leg movements and of body movements (cf. Stier, 1930), and the tonus of the skele-

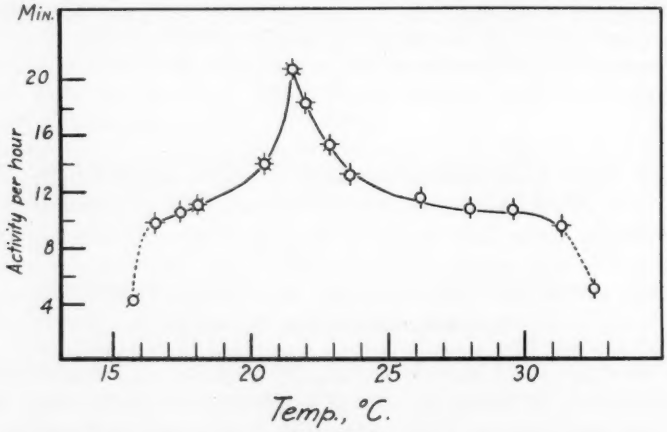


FIGURE 1

Calculated total "spontaneous" activity per hour of two-day old albino mice as a function of temperature. Ordinate: the product of the average duration of single periods of activity times the number of periods per hour. Abscissa: body temperature, °C. The observations are plotted as averages of data from several litter mates; each spoke on the circle refers to the number of animals included in each average.

tal muscles, increases with temperature. The additional heat contributed by these sources should likewise change with temperature. The heat production, however, as indicated by thermopile measurements of the body temperatures of young albino mice (Pincus, Sterne and Enzmann, 1933) of the same strain used by Stier (1930), increases as the body temperature is raised from 15° to 25°C.; and decreases progressively from 25° to 35°C. The diminution of heat production beyond 25° can be possibly explained by the decrease in duration of activity periods with temperature and by the peculiar relation of the frequency of occurrence of activity

periods to temperature. The number of activity periods per hour increases as the body temperature is progressively raised from 16° to 24°C. In the neighborhood of 25° a sudden diminution in the number of periods occurs (34 at 24°; 20 at 26°). Thereafter, the frequency again increases as the body temperature is increased (cf. Stier, 1930). It may be found in subsequent experiments that the "intensity" of muscular activity follows a similar relation to temperature; or that it steadily increases from 15 to 35. The net effect of the additional heat contributed by the "intensity" of spontaneous movements, no matter what the type of temperature relationship, appears to be of minor importance in accounting for the heat production above 25°. It has already been found (Pincus, Sterne and Enzmann, 1933) that the difference between internal and external temperature rises to a maximum at 25° and then falls off at higher body temperatures. Below 25° the heat contributed by the intensity of movements and by other heat producing processes (e.g., cellular metabolism) may account for the difference in location of the maxima in the graphs by Pincus, Sterne and Enzmann (1933) at about 25° and the peak in figure 1 at about 22°. A quantity referring to the intensity of muscular activity should be therefore included in the product—duration of activity periods times frequency of occurrence— in order to account entirely for the rate of heat production by the active muscles. A measure of this quantity will be sought in subsequent experiments.

The data of Pflüger (Starling, 1930) on the O_2 consumption of rabbits in the range of internal temperatures from 22° to 38.8°C. yield graphs quite similar to figure 1, the peak occurring at 37.5°C. In Martin's experiments (cf. Starling, 1930) on the CO_2 production of rabbits and measurements of the O_2 consumption of mice by Benedict and Fox (1933), only the external temperature was recorded and so the data cannot be used for our present analysis. More experiments in which spontaneous activity is measured over a wide range of internal temperatures are needed to complete the study of temperature regulation.

Heat regulation by the adult mammal can be upset by injuring the *corpus striatum* (Barbour, 1921). As yet it is not known whether the "center" controlling the periodicity of movements is identical with the heat regulatory center, or whether it is closely associated with it. Tracy (1926) has pointed out that the center controlling "spontaneous" movements may be closely associated with the respiratory center.

On the basis of the information contained in figure 1 and the findings of Pincus, Sterne and Enzmann (1933), it is suggested that in the range of body temperatures above 25° (approximately above 23.7°C., air temperature for mice of these ages) the two-day old mouse already exhibits a kind of regulation of its body temperature. If, for instance, a young albino mouse kept in a nest at, say, 30° is moved out into a room at 25° there

occurs (after one hour) a 13.5 per cent increase in the product: duration times number of activity periods per hour; if, however, an animal is taken from the same nest to an environmental temperature of 32°C. there occurs a 32.7 per cent decrease in the total amount of muscular movements (as measured by the units plotted in figure 1). There is thus a tendency to produce more heat at temperatures below that of the nest and less heat at temperatures higher than that of the nest. The extra heat produced at lower temperatures is of course rather rapidly lost in these young furless individuals. However, as the animal becomes older, thicker layers of insulating materials are acquired, heat is lost less rapidly when the temperature of the environment falls, and consequently the fluctuations of internal temperature become narrowed in their extent. So in conjunction with the other heat-producing processes and the heat-conserving and -dissipating mechanisms, the body temperature of the adult animal is held constant at some fixed point.

Experiments now in progress with older mice of the same inbred strain (the Bagg albinos) used in the above experiments (Stier, 1930) will test whether a similar mechanism controlling the heat production at different body temperatures is retained in adult animals. If the mechanism outlined above for young mice exists in adult animals one can then understand why a decrease in the body temperature of the so-called homeothermic mouse brings about an increase in the amount of movement of the animal and so an increase in heat production; and an increase in the external temperature brings about a decrease in the amount of bodily activity.

It is important in studying the nature of temperature regulation in mammals to ascertain the magnitude of the internal temperature maintained over long periods of time and the actual heat production at many steps in a wide range of body temperatures, studying at the same time the development of the mechanisms which dissipate and retain the heat contributed by muscular activities and by other heat-producing processes. This procedure is being followed in experiments by Dr. G. Pincus and the writer upon several strains of inbred mice at various ages after birth.

Summary.—The total amount of "spontaneous" muscular activity of two-day old albino mice increases as the animal's internal temperature is raised from 16° to 22°C. and then steadily decreases at successive higher temperatures. Similar graphs have been obtained for the difference between external and internal temperatures of young albino mice of the same strain in the range of body temperatures, 15° to 35°C. (cf. Pincus, Sterne and Enzmann, 1933). The maximum difference occurs at about 25°C.

It is suggested that there is exhibited a kind of regulation of temperature by the young mouse in the range of internal temperatures above 25°. In this upper temperature range a decrease in air temperature results in an increase in the amount of muscular activity and an increase in the heat

production; an increase in the temperature of the air is followed by a decrease in the amount of muscular activity and a decrease in the production of heat.

* The work in this paper was supported by a grant from the Milton Fund for Research, Harvard University.

Barbour, H. G., *Physiol. Rev.*, **1**, 295 (1921).

Benedict, F. G., and Fox, E. L., *Pflüger's Archiv. f. d. ges. Physiol.*, **231**, 455 (1933).

Pincus, G., Sterne, G. DeRoo, and Enzmann, E., *Proc. Nat. Acad. Sci.*, **19**, 729–733 (1933).

Starling, E. H., *Principles of Human Physiology*, Lea and Febiger, 5th Edition, 943 (1930).

Stier, T. J. B., *J. Gen. Psych.*, **4**, 67 (1930).

Stier, T. J. B., and Pincus, G., *J. Gen. Physiol.*, **11**, 349 (1927–28).

Tracy, H. C., *J. Comp. Neur.*, **40**, 253 (1926).

---

*THE DEVELOPMENT OF TEMPERATURE REGULATION IN THE MOUSE**

BY G. PINCUS, G. DEROO STERNE AND E. ENZMANN

LABORATORY OF GENERAL PHYSIOLOGY, HARVARD UNIVERSITY

Communicated June 9, 1933

In experiments dealing with the temperature relations of certain vital processes in young mice (Stier and Pincus, 1927–28; Stier, 1930; Pincus, 1931) it was noted that the body temperature of even one-day old mice was ordinarily not quite that of the environment, despite their complete lack of fur and their small size, but, over the range 15° to 35°C., somewhat above that of the environment. The difference between external and internal temperatures seemed to be roughly a function of the age of the animal, such that the older the mouse the greater the difference. It was thought, therefore, that by measuring this difference in a series of mice from birth to maturity a quantitative measure might thereby be had of the relation of the ability for temperature regulation to age. The simple procedure would seem to be to employ one or more standard constant external temperatures and to ascertain the difference between these temperatures and the internal temperature of mice of various ages, and then by plotting this difference against age a presumably rising curve would be had indicative of the nature of the age-regulatory relation (cf. Baldwin and Kendeigh, 1932, p. 108). Such a procedure was found to be quite inadequate, however, because of two definite difficulties.

The first difficulty involves the exact estimation of the internal temperature of a partially homeothermic mouse kept at a constant external temperature. The technique employed, described in detail by Stier and Pincus

(1927–28), involves the measurement of rectal temperatures by the use of a thermopile.[1] The rectal temperature of a young mouse taken from the nest and placed in a thermostat kept at a temperature of, say, 16°C. decreases fairly regularly with time. The shapes of the cooling curves of live mice differ markedly from those for dead mice of the same age. This is illustrated in figure 1, where the cooling curves of a live and of a dead mouse of approximately the same age and weight, and kept at approximately the same temperature, are plotted logarithmically. The dead mouse exhibits a typical Newtonian cooling curve, as demonstrated by the

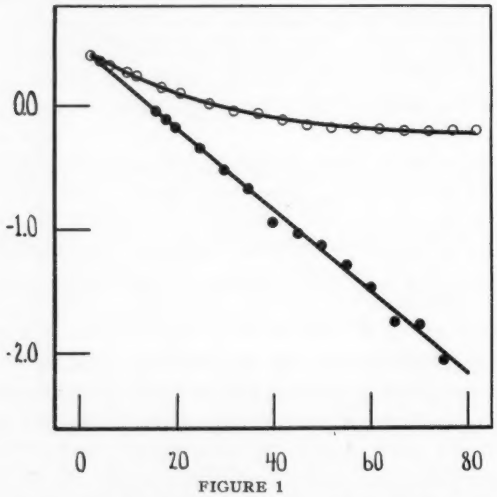


FIGURE 1

Comparison of the cooling curves of live and dead mice. Ordinate: logarithm of the difference between internal and external temperatures. Abscissa: time in minutes. Open circles = 5-day old mouse at 15°.32 ± 0°.02. Closed circles, 6-day old dead mouse at 15°.95. Note difference in rate of cooling and in the shape of the cooling curves. See text.

linear relation between elapsed time and the logarithm of difference between external and internal temperature. For purposes of comparison the cooling curve of the dead mouse has been extrapolated to intersect the curve for the live mouse at a point where the difference between external and internal temperatures in both live and dead mice is 2°.52C. (0.4014 in logarithmic units), and the comparative rates of cooling may be adjudged by the slopes of the respective lines. The rate of cooling is much slower in the live mouse, nor does it follow Newton's law. It is as if some internal mechanism had put a "brake" upon the cooling process. It is conceivable

that a measurement of this "braking" would give a measurement of the way the temperature-regulating mechanism responds to temperature changes.

When a dead mouse is warmed, the rate of attainment of the higher temperature is also logarithmic. When a live mouse is placed at a higher temperature, however, quite a different situation is observed. The rate at which the live mouse attains the higher temperature is now much greater than the rate of warming of a dead mouse of the same age and weight, and the live mouse, within a certain temperature range at least, extremely soon attains a body temperature above that of the environment. So rapidly does the body temperature of a live mouse rise when the animal is placed at a higher temperature that it is often difficult to obtain readings for the

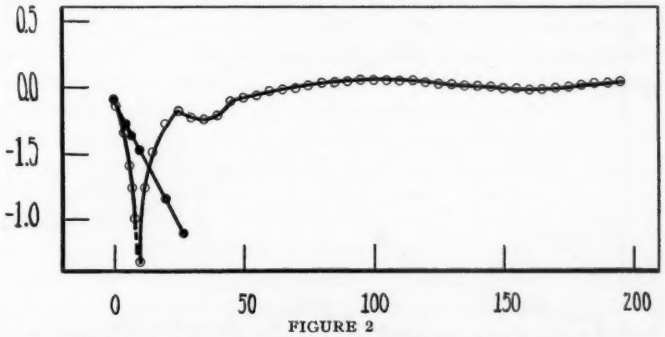


FIGURE 2

Comparison of the warming curves of live and dead mice. Ordinate and abscissa as in figure 2. Open circles = 2-day old live mouse at 32°.81 ± 0°.02. Closed circles = 2-day old dead mouse at 30°.70 ± 0°.02. Note early rapid attainment of external temperature by live mouse and subsequent rise of internal temperature. Observe also periodic fluctuations in internal temperature of live mouse.

rate of warming. The course of events is illustrated in figure 2. The first five points on the curve for the live mouse represent the period when the body temperature of the animal is below that of the environment. Thereafter the body temperature rises above that of the environment, but not continuously or smoothly. The peaks and valleys in the curve represent periods of rise and fall in the body temperature. We continually encounter such evidences of rhythmic change in internal temperature, and our estimation of the "normal" level of body temperature is made difficult because of the presence of these rhythms. Such rhythms are, of course, never encountered in dead animals.

If we discount the spontaneous rhythmic changes, and proceed to estimate the "normal" body temperature of a series of young mice at various constant external temperatures, our second and most serious difficulty

is encountered. This is revealed when we ascertain the difference between external and internal temperature as a function of temperature. Such data are plotted in figure 3 for mice of three different ages. It is apparent that the curves relating temperature difference to temperature are cusped in each case. Over the range 25°± to 35°± this difference decreases (cf. Stier, 1933). Over the lower temperature range we observe the ordinary relation of chemical processes to temperature (the temperature coefficient is positive and high); over the higher range we have a negative temperature coefficient. The latter relation is what one would expect in an animal

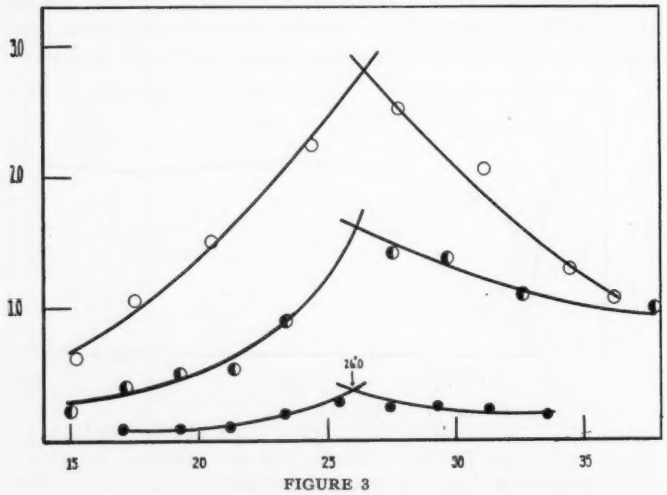


FIGURE 3

Calculated "normal" difference between internal and external temperatures for live mice of various ages as a function of temperature. Ordinate: difference between internal and external temperatures in °C. Abscissa: internal temperatures in °C. Closed circles = 2-day old mouse. Half circles = 8-day old mouse. Open circles = 10-day old mouse. See text.

capable of temperature regulation, i.e., the higher the temperature the less the difference between external and internal temperature. It will be noted, too, that the slope of the curves is apparently steeper over the lower temperature range. The rising curve occurring over the lower temperature range we have designated the "metabolic" response, the second curve the "regulatory" response. Stier (1930) has noted a similar critical temperature at 25°C. concerned with certain aspects of spontaneous activity of one-day old mice. He employed mice of the same inbred strain (Bagg-Little albinos).

It becomes difficult to ascertain the development of the temperature-regulating mechanism with age, because the most elementary comparison

would require knowing whether the constant external temperature used results in the attainment of an internal level proper to the "metabolic" phase or to the "regulatory" phase. Whether a given level falls into one or the other category is apparently conditioned not merely by the external temperature but also by the internal temperature. Furthermore, we have some evidence that in older animals the "regulatory" response may occur at temperatures below 25°C. Experiments now in progress are being conducted to ascertain exactly the nature of these two phases.

The data of figure 3 roughly illustrate that with increasing age the difference between external and internal temperature increases in both the "metabolic" and the "regulatory" phases. The oldest mouse examined in the present series was 24 days old, and even at that age perfect regulation, i.e., the ability to maintain the body temperature at a constant high level (e.g., 36°8C.) at any outer temperature, was not observed. The body temperature of adult mice of the same strain is normally 36°8±, but it is possible to lower this temperature by various means. Details of experiments with adult mice will be given elsewhere.

Summary.—The rate of cooling of living mice aged 2 to 24 days is always slower than that of dead mice of the same age and weight. Conversely, the rate of warming of live mice is always faster than that of dead mice of the same age and weight. At any constant external temperature over the range 15° to 35°C. the rectal temperature of living mice is always somewhat above the outer temperature. The equilibrium difference between internal and external temperature is not constant, but fluctuates rhythmically about a mean. When the difference between external and internal temperature, for mice of the ages used, is plotted as a function of temperature over the range 15° to 35°C., it is found that this difference increases to a maximum at 25°C.± and then falls. The ascending limb of the curve has been designated the "metabolic" phase, the descending limb the "regulatory" phase. No exact description of the relation of the development of temperature-regulating ability to age can be given without the resolution of the rôle of these two phases at all ages.

* The work summarized in this paper was in part supported by the Milton Research Fund, Harvard University.

[1] In the experiments described in this paper the mice used were albinos of the 42nd to 45th brother-sister generation of the Bagg-Little strain.

Baldwin, S. P., and Kendeigh, S. C., *Physiology of the Temperature of Birds*, Cleveland (1932).

Pincus, G., *J. Gen. Physiol.*, **14**, 421 (1931).

Stier, T. J. B., *J. Gen. Psychol.*, **4**, 67 (1930).

Stier, T. J. B., 1933. In press.

Stier, T. J. B., and Pincus, G., *J. Gen. Physiol.*, **11**, 349 (1927–28).

THE INFLUENCE OF USE OR FREQUENCY OF OCCURRENCE UPON THE STRENGTH OF MENTAL CONNECTIONS

BY EDWARD L. THORNDIKE

INSTITUTE OF EDUCATIONAL RESEARCH, TEACHERS COLLEGE

Read before the Academy, Tuesday, April 25, 1933

Until recently the doctrine in mental dynamics which has been accepted most widely and used oftenest as an explanation of learning is the doctrine that the mere occurrence of a connection strengthens it. It has seemed almost axiomatic, but almost all the evidence showing that use strengthens connections is contaminated by the presence of satisfying after-effects, so that it is possible to maintain with some reason that the occurrence or use of a connection does nothing but give an opportunity for other forces to act upon it. The Gestaltists have also challenged the doctrine on general grounds.

In *The Fundamentals of Learning* evidence was reported showing that the mere occurrence of a connection does strengthen it, in the shape of the strengthening found where after-effects of any sort were at a minimum, and also in the shape of the strengthening of connections which were punished. The latter evidence was not treated fully or critically in that volume. It has been extended by Lorge and the writer, and will be the topic of the present report.

In experiments with human multiple-choice learning with reward and punishment, a connection which occurs and is punished often is strengthened more by the occurring than it is weakened by the punishment. Evidence of this has been presented by Thorndike and others ['32, p. 108 ff. and p. 276 ff.] and by Lorge ['33]. We shall determine the extent to which this happens and the amount of the balance of strengthening, and observe the characteristics of individuals in respect of it. Experiments made to measure the "spread" or "scatter" of the influence of rewards and the influence of the relevance of rewards provide the necessary data.[1]

Consider first the facts from an experiment (Q) in learning one response out of six for each of 960 situations, the response being the numbers 1 to 6 and the situations 960 words, divided into 24 sets of 40. A word was read by the experimenter; the subject said 1 or 2 or 3 or 4 or 5 or 6; the experimenter said "Right" or "Wrong," and then read the next word, and so on through a series of 40, and through the same series four times more. Each "Right" signified a money payment which was of real importance to the learner. There were 24 subjects.

The influence of one rewarded occurrence in Experiment Q is such that the connection which occurs and is rewarded occurs again in the following trial or round in 4029 out of 9751 cases, or in 43.6 per cent.[2]

The influence of one punished occurrence is such that the connection which occurs and is punished occurs in the following trial or round in 9868 out of 43,915 cases, or in 22.5 per cent.

If we assume that, apart from the occurrences, rewards and punishments, the subject would at any occasion be as likely to choose any one number as any other, the rewarded occurrence has increased the probability of reoccurrence by 0.269 (from 0.167 to 0.436) and the punished occurrence has increased it by 0.058 (from 0.167 to 0.225).

Some of the latter increase is, however, probably due to the "spread" or "scatter" action of rewards, and some of it may be due to favoritism in the choice of numbers. We will consider the latter first.

The percentage of repetitions which there would have been if the punishment and occurrence had had a net influence of zero would be 100 divided by the number of different responses (here 6), if the subject was as likely to select one response as another. But if he favored certain responses it would not. In any trial (call it II) the probability of repetition of a response made in the preceding trial (call it I) apart from all influence from the learning itself, is aI aII + bI bII + cI cII.....nI nII where aI, bI, cI, etc., are the probabilities of occurrence in trial I and aII, bII, cII, etc., are the probabilities of occurrence in trial II. Favoritism for certain numbers can lower the probability of repetition to 0 or raise it to 1.00. For example, in a case where the subject must respond by any number from 1 to 6, we could have such cases as *A*, *B*, *C* and *D* below (the occurrences total 1000 in each trial).

Case A:

Trial I	1,	400;	2,	300;	3,	200;	4,	100;	5,	0;	6,	0.
Trial II	1,	0;	2,	0;	3,	0;	4.	0;	5,	400;	6,	600.

Probability of repetition, 0.000.

Case B:

Trial I	1,	1000;	2,	0;	3,	0;	4,	0;	5,	0;	6,	0.
Trial II	1,	1000;	2,	0;	3,	0;	4,	0;	5,	0;	6,	0.

Probability of repetition, 1.00.

Case C:

Trial I	1,	400;	2,	300;	3,	200;	4,	100;	5,	0;	6,	0.
Trial II	1,	400;	2,	300;	3,	200;	4,	100;	5,	0;	6,	0.

Probability of repetition, 0.300.

Case D:

Trial I	1,	350;	2,	250;	3,	200;	4,	100;	5,	50;	6,	50.
Trial II	1,	360;	2,	240;	3,	180;	4,	100;	5,	60;	6,	60.

Probability of repetition, 0.238.

Favoritism of a similar sort in both trials will raise the probability of repetition above 0.1667. Reversal of favoritism in the second trial will lower it below 0.1667.

We therefore proceed to investigate the favoritism, if any, shown by each of the twenty-four subjects in choosing among 1, 2, 3, 4, 5 and 6 as responses. The responses of each are tabulated in successive sets of 40 corresponding to the successive trials of the various series of words, then in successive sets of 200, each corresponding to the 200 responses for the five trials of one series, and finally by successive sets of 400.[3] First, each record is inspected to detect any tendencies to reverse favoritism in any two consecutive trials. No such were found.

Secondly, each record is inspected to detect any systematic shifts in favoritism (such as a steady change from favoring 1 and 2 much to favoring them little, or from favoring 2, 3, 4 and 5 to favoring them little or vice versa). There were few such shifts.

Thirdly, each record is inspected to detect any clear cases of short-lived changes in favoritism. It is, of course, hard to draw the line between such and chance variations in choices.

If there is no evidence of systematic shifts, or of genuine short-lived tendencies, the average influence of favoritism operating at any one time is estimated as follows: $a^2 + b^2 + c^2 + d^2 + e^2 + f^2 / n^2$ is computed for each successive 400 (or for so many 400's as we have) and the average of the determinations is taken. This permits undetected genuine short-lived tendencies to favoritism to count somewhat, but still does not interpret chance variations in choice as genuine favoritisms. If there are any special shifts, the 400's are so chosen as not to swamp the influence of the favoritism, and are checked by using sets of 200 each.

As a result of much labor expended in the ways stated, a probability of repetition of the same number in the responses to the same word in two successive trials has been estimated for each of the 24 subjects.[4] They range from 0.170 to 0.220, twenty being between 0.174 and 0.205. These are shown in table 1 together with the excess frequencies of repetition for punished occurrences at various degrees of remoteness from any rewarded connection.

Even at a remoteness of five or more steps from any rewarded connection the frequency of repetition in the next trial for one punished occurrence of a connection is clearly in excess of what would be expected by chance plus number favoritism. It is + in 19 cases out of 24, and the median (16 permilles) is about three times its probable error. A punished connection gains more strength by occurring once than it loses by being punished once.

Why this should be so is an interesting problem. Some facts and possibilities are as follows: (1) Lorge and Thorndike ['33] have shown that in man a connection left ambiguous and neutral by having no announcement or reward or punishment of any sort attached to it gains still more. (2) The net gains shown in table 1 may then be reductions from still

larger gains due to the occurrences pure and simple. (Lorge does not, however, regard this as probable.) (3) The strengthening by the mere occurrence may be reduced not at all by the punishment in these experi-

TABLE 1

THE STRENGTHENING DUE TO ONE (FIRST) OCCURRENCE OF A PUNISHED CONNECTION. EXPERIMENT Q.

	EXCESS OVER THE EXPECTED FREQUENCY FOR THE ACTUAL FREQUENCIES OF REPETITION IN THE FOLLOWING TRIAL FOR FIRST OCCURRENCES OF PUNISHED CONNECTIONS. PERMILLES							
	BY CHANCE + FAVORITISM FOR NUMBERS	ALL PUNISHED CONNECTIONS	BETWEEN TWO REWARDED CONNECTIONS	NEXT TO ONE REWARDED CONNECTION	TWO STEPS REMOVED FROM EACH OF TWO REWARDED CONNECTIONS	TWO STEPS REMOVED FROM ONE REWARDED CONNECTION	THREE OR FOUR STEPS REMOVED FROM ONE OR TWO REWARDED CONNECTIONS	FIVE OR MORE STEPS REMOVED FROM ANY REWARDED CONNECTION
	1	2	3	4	5	6	7	8
1	200	109	145	107	157	107	119	091
2	174	033	008	038	063	053	039	026
3	170	027	071	057	008	009	024	002
4	189	−003	071	005	081	−051	−011	005
5	185	010	041	017	093	019	011	003
6	196	007	057	013	035	025	−023	024
7	210	030	167	036	−130	026	030	−017
8	199	031	126	023	−045	025	017	050
9	175	064	075	075	044	081	082	019
10	185	071	081	073	194	090	070	077
11	174	020	072	050	039	−006	−031	−036
12	208	035	109	045	023	049	026	−009
13	180	039	146	044	106	−005	027	030
14	180	025	107	035	120	037	023	002
15	180	050	060	039	−013	046	040	069
16	174	045	022	081	159	023	038	042
17	177	004	058	055	109	−001	−027	−020
18	182	042	032	079	100	012	057	004
19	205	014	032	022	045	007	028	−019
20	189	111	061	094	144	078	144	107
21	183	048	117	075	140	039	035	011
22	190	052	174	081	041	013	038	042
23	220	033	018	014	165	049	044	042
24	180	008	111	000	062	−004	−008	013
Median	184	032	071.5	044.5	072	025	029	016
Average	188	038	082	048	073	030	033	023

ments, so that the 016 is the strengthening due to one occurrence pure and simple. (4) The spread or scatter effect may extend out further than past information has led us to expect. (5) There may be a strengthening due to the satisfyingness of making any choice and response, which

is reduced by the annoyingness of finding it to be wrong. (6) Certain words may more or less persistently suggest certain numbers to certain subjects by reason of their meanings or associations. We shall return to this problem after reporting the facts for other experiments.

Consider now the facts for Experiment Hl and Ha (described in "An Experimental Study of Rewards," Section II) in which the learning was to connect a number from 1 to 10 with each of 480 words (Hl) and 160 words (Ha). Two readings of each series of 40 words and the right numbers to the subject before his first trial with that series put the chance for a right choice in trial 1 well above 0.10, but that does not concern us here. The percentage of repetition in the trial following a first occurrence that was punished was far in excess of 0.10, being over 0.24, and was 0.21 when the punished connection was five steps or more removed from any reward. We have to discover how much of this excess was due to number favoritism.

Experiment Hl really comprises ten different sub-experiments, each involving the responses to 40 words in each of four trials. What we really need to know, in order to allow perfectly for the influence of favoritism for numbers, is the status of favoritism before each response of each of the sub-experiments, or, since that is impracticable, at the beginning of each trial of each sub-experiment. That is, we should like to know the nature and amount of favoritism at any stage in the learning, regardless of how it was caused, so as to allow for its action point by point, or at least trial by trial, in the course of the learning. But even the favoritism trial by trial would be exceedingly difficult to estimate. The size of the samples (40) is too small for reliable measurements, and we should have to work out correction factors. Until the need for allowance trial by trial becomes acute nobody will, or should, undertake the labor. We may, however, reasonably undertake the labor of measuring the favoritism in the first two sub-experiments, the second two sub-experiments, etc., since that will give us N's of 320, which should not increase the obtained favoritisms greatly above the real tendencies which they sample. We have done this with the facts for the first 960 choices of Experiment Hl, and the 640 choices of Experiment Ha, with the results shown in table 2.

If the tendency of a person in the 40 choices of any trial in Hl was to favor numbers as he did in the entire 960, the probability that, apart from any other influences than his number favoritism, he would give the same number for any given word in any two successive trials would be 0.113 for Bra, 0.104 for Bre and 0.117, 0.105, 0.110, 0.103, 0.115, 0.120, 0.111 and 0.106 for the other subjects in order.

In so far as short-lived habits or systematic shifts of favoritism existed these figures are too low. They are not much too low. The division of the 960 into three 320's in order of occurrence and the computation of the average of the same probabilities from the three 320's gives results

TABLE 2

FAVORITISM FOR NUMBERS. THE FREQUENCIES WITH WHICH BRA, BRE, ETC., USED 1, 2, 3, ...10 IN SUB-EXPERIMENTS 1 AND 2, 3 AND 4, AND 5 AND 6 OF EXP. HL AND IN 1 AND 4 AND 7 AND 10 OF EXP. HA

	BRA					BRE					BL				
	1,2	3,4	5,6	1,4	7,10	1,2	3,4	5,6	1,4	7,10	1,2	3,4	5,6	1,4	7,10
1	48	35	31	27	25	27	47	38	33	33	31	24	16	27	15
2	33	47	25	36	37	23	38	28	10	33	35	51	31	46	31
3	37	40	35	26	30	47	32	31	40	30	42	41	40	32	32
4	60	48	56	54	42	39	41	50	32	36	16	8	27	35	46
5	18	30	28	33	32	35	19	21	29	36	36	22	48	53	55
6	37	41	37	37	49	17	44	41	38	24	33	41	28	42	39
7	30	19	33	17	25	38	34	40	25	59	44	22	37	20	9
8	14	16	21	22	29	34	34	18	55	33	35	33	32	26	19
9	43	30	29	31	32	35	14	25	14	22	31	34	19	38	49
10	0	14	24	37	19	25	17	28	0	14	17	44	42	1	25

	A. BU.					L.					M.				
	1,2	3,4	5,6	1,4	7,10	1,2	3,4	5,6	1,4	7,10	1,2	3,4	5,6	1,4	7,10
1	23	28	31	39	28	7	11	26	24	36	42	42	26	22	38
2	65	40	51	47	56	28	32	42	31	26	14	22	41	28	50
3	41	39	40	37	36	53	38	41	53	34	46	16	28	28	27
4	50	54	46	45	54	46	48	43	30	38	30	48	50	61	25
5	13	12	5	33	13	37	29	26	61	41	41	27	18	28	48
6	8	21	26	22	27	55	40	30	34	51	26	29	35	52	30
7	29	32	27	23	36	14	31	22	29	17	35	34	32	42	43
8	39	42	49	38	21	28	21	40	21	24	27	37	27	16	21
9	23	19	27	26	36	34	57	25	30	24	23	31	20	21	16
10	29	33	18	10	13	18	13	25	7	29	36	34	43	22	22

	PEI					RO.					E. T.				
	1,2	3,4	5,6	1,4	7,10	1,2	3,4	5,6	1,4	7,10	1,2	3,4	5,6	1,4	7,10
1	29	30	30	14	19	29	41	46	29	39	30	18	23	22	25
2	39	30	48	11	22	58	57	51	64	46	27	32	34	31	25
3	46	38	25	31	33	29	13	13	15	9	45	27	42	39	46
4	22	43	41	56	55	21	17	10	22	14	39	51	44	31	25
5	63	56	61	30	63	37	35	45	51	54	29	36	45	34	35
6	41	46	31	33	31	64	27	43	40	34	53	57	52	55	33
7	28	22	17	53	48	23	54	48	39	62	26	19	20	20	43
8	11	13	25	45	25	21	31	17	19	5	20	20	23	28	37
9	14	14	26	33	10	14	5	1	19	16	26	37	17	45	36
10	27	18	16	14	14	24	40	46	19	41	25	23	20	15	15

	R. T.				
	1,2	3,4	5,6	1,4	7,10
1	30	30	36	16	20
2	23	29	47	44	42
3	43	39	33	45	41
4	40	42	32	35	38
5	16	22	31	34	33
6	47	41	36	38	39
7	26	15	18	38	27
8	39	46	38	19	25
9	34	32	32	42	32
10	22	24	17	9	23

only 0.0044 higher. Inspection of the choices by successive 160's shows little or nothing striking in the way of short-lived habits or systematic shifts. Moreover, these subjects in the similar experiment (Ha) over a week before Experiment Hl, showed much the same sorts of favoritisms as they showed in Hl.

We have computed the sums of the differences between the number of choices of the same number in the various 320's. For neighboring 320's the sums are 922, 856 and 998 (averaging 925). For the first 320 and third 320 of the 960, the sum is 1029. For any 320 of Hl and any 320 of Ha the sums are 1281, 1162, 1171, 1109, 1113 and 997 (averaging 1139). We shall therefore not be far wrong if we use the average of the separate probabilities for the five 320's of Experiment Ha and Hl as the base from which to measure the influence of one punished occurrence in Experiments Ha and Hl.

Table 3 presents, for each subject in Experiment Ha and Hl, observed frequencies of repetition in the following trial of a punished connection which at its first occurrence was five or more steps removed from any rewarded connection, along with this base (and also the five probabilities of which it is the average). There is in all subjects an excess of repetition above that explicable by favoritism for numbers. The average excess (092) is $6^1/_2$ times its probable error. The individual differences are large, as they were in Experiment Q.

TABLE 3

THE OBSERVED FREQUENCY OF REPETITION IN THE FOLLOWING TRIAL OF A PUNISHED CONNECTION FIVE OR MORE STEPS REMOVED FROM ANY REWARDED CONNECTION AT ITS FIRST OCCURRENCE (OBS.), AND THE APPROXIMATE FREQUENCY OF REPETITION TO BE EXPECTED FROM NUMBER FAVORITISM ALONE (FAV.), IN PERMILLES. ALSO THE PROBABILITIES FOR EACH COLUMN OF TABLE 2. THE DATA ARE FROM EXP. HL AND HA

	OBS.	FAV.	EXCESS OF OBS. OVER FAV.	EXP. Hl 1,2	EXP. Hl 3,4	EXP. Hl 5,6	EXP. Ha 1,4	EXP. Ha 7,10
Bra	132	113	019	127	114	109	109	107
Bre	124	111	013	107	112	109	116	112
Bl	298	114	184	107	115	109	119	121
A. Bu	171	118	053	126	114	119	112	119
L	167	116	051	123	119	106	121	109
M	298	112	186	109	108	109	119	113
P	235	121	114	122	116	118	122	128
R	277	128	149	124	125	133	122	135
E. T.	237	112	125	109	116	115	113	108
R. T.	135	109	026	109	109	107	114	106
Median			084					
Average			092 (P.E. = 014)					

In the case of Experiment F (in which the choice was of a number from 1 to 10) the individual records of actual repetitions of punished connections remote from a reward were from too few cases to deserve computation.

The totals for the group of twenty subjects were: 605 first occurrences five or more steps removed from any reward, of which 74 or 12.2% were repeated in the following trial. The probability of a repetition apart from occurrences, rewards and punishments computed from the frequency of choices in the first 4 trials varies from 0.109 to 0.141, averaging 0.120. When computed from the last two trials, it varies from 0.102 to 0.231, averaging 0.140. The two medians are 0.1175 and 0.136. The frequency of repetition by chance plus number favoritism may then be set as approximately 0.130. There thus appears to have been no balance of strengthening effect from an occurrence of a punished connection in this experiment.

For Experiment P (described in Lorge and Thorndike, p. 374) table 4 gives the facts concerning the expected frequency of repetition for each individual by chance plus his number favoritisms and concerning the actual repetition for punished connections 5 or more steps removed from any reward.[5] The number of observations for an individual is small (about 85 in Part A and about 45 in Part B) so that the per cents in column 6 and the differences in column 7 are not individually reliable. But the

TABLE 4

THE STRENGTHENING DUE TO ONE (FIRST) OCCURRENCE OF A PUNISHED CONNECTION IN EXPERIMENT P. IN PERMILLES.

	EXPECTED FREQUENCY OF REPETITION BY CHANCE PLUS NUMBER FAVORITISM					ACTUAL REPETITION OF PUNISHED CONNECTIONS 5 OR MORE STEPS FROM A REWARDED CONNECTION	ACTUAL MINUS EXPECTED BY CHANCE PLUS NUMBER FAVORITISM 6–5
	FIRST FOUR $n = 512$	LAST FOUR $n = 508$	FIRST TWO $n = 256$	TWO WITH MOST FAVORITISM $n = 252$	ESTIMATE FOR EXPERIMENT AS A WHOLE		
PART A	1	2	3	4	5	6	7
B	290	258	286	304	280	353	073
H	252	254	250	359	265	337	072
L	264	279	265	282	270	333	063
M	256	256	258	275	260	567	307
Sc	260	251	254	271	260	402	142
Sh	260	257	263	264	260	216	−044
T	255	251	255	271	260	628	368
Ca. W.	286	286	295	304	290	254	−036
Ch. W.	261	255	257	266	260	255	−005
PART B	$n = 352$	$n = 352$	$n = 176$	$n = 176$			
Bel	254	255	263	287	260	452	192
D	251	254	252	261	255	209	−046
La	258	250	266	295	260	429	169
Al	304	442	330	487	370	400	030
Fl	257	261	259	282	260	375	115
Ri	256	255	257	262	255	295	040
Ro	287	252	286	296	275	408	133
J. S.	252	259	256	302	265	500	235
T. S.	259	259	264	267	260	385	125
We	279	264	305	305	270	224	−046

general results (median excess over what chance and number favoritism would give, 073) is about four times its probable error.

A similar treatment of the records in Experiment M results in the facts of table 5. The measurements for any one individual are from too few cases to be reliable, but the average (0.037 with a P. E. of 0.016) indicates that a punished occurrence has a strengthening influence upon the connection in question in this experiment as in Experiments Q, H and P.

TABLE 5

THE STRENGTHENING DUE TO ONE PUNISHED OCCURRENCE IN EXP. M, COMPUTED BY $P_{Act}-P_{AV}$. P_A, P_B, P_C AND P_D ARE THE PROBABILITIES OF A REOCURRENCE OF THE SAME RESPONSE IN TWO CONSECUTIVE TRIALS IN SUB-EXPERIMENTS *A*, *B*, *C* AND *D*, APART FROM ANY INFLUENCE OF OCCURRENCES, REWARD AND PUNISHMENTS. P_{AV} IS THE AVERAGE OF THESE FOUR PROBABILITIES, P_{Act} IS THE ACTUAL FREQUENCY OF REPETITIONS IN THE FOLLOWING TRIAL OF PUNISHED CONNECTIONS FIVE OR MORE STEPS REMOVED FROM ANY REWARD (ONLY FIRST OCCURRENCES BEING USED). *n* = THE NUMBER OF CASES USED IN DETERMINING P_{Act}

SUBJECT	P_A	P_B	P_C	P_D	P_{AV}	P_{Act}	*n*	STRENGTHENING DUE TO ONE PUNISHED FIRST OCCURRENCE
A	0.171	0.226	0.307	0.218	0.231	0.284	67	0.053
B_1	0.179	0.170	0.170	0.187	0.177	0.221	68	0.049
B_2	0.195	0.257	0.178	0.180	0.203	0.085	82	−0.118
B_3	0.245	0.208	0.235	0.265	0.238	0.346	78	0.108
L	0.173	0.183	0.192	0.172	0.180	0.176	85	0.004
M	0.181	0.198	0.180	0.174	0.183	0.226	62	0.043
O	0.353	0.350	0.454	0.297	0.364	0.477	44	0.113
N	0.186	0.209	0.182	0.189	0.192	0.240	129	0.048

For the 20 subjects of Experiment J the facts were as shown in table 6.[6] There is an excess of repetition due to one punished occurrence 5 or more steps removed from any rewarded connection in every individual. The median is 98 permilles, which is about 10 times its probable error.

For the 8 subjects of Experiment L, the median excess was 46 permilles.[6]

In experiments upon the influence of varying amounts of money reward,[7] Rock found the percentage of repetition of punished connections to be 33.5 where pure chance would give 25.0 and where chance plus position favoritism would give (according to the facts in a similar experiment) about 27.0. Even after a liberal allowance for the influence of the spread or scatter, there will remain a substantial amount of strengthening due to the mere occurrence of the connection.

With the single exception of Experiment F, a punished connection gains more strength from its first occurrence than it loses by being punished. In my opinion, if the situation vanishes so promptly that the person is not led by the punishment to change his response to it, i.e., to create some second and different connection, and if he is immediately engaged in some new task so that he does not create some second and different connection

TABLE 6

THE STRENGTHENING DUE TO ONE OCCURRENCE OF A PUNISHED CONNECTION 5 OR MORE STEPS REMOVED FROM ANY REWARD. EXPERIMENT J. IN PERMILLES.

SUBJECT	REPETITION DUE TO CHANCE AND NUMBER FAVORITISM				ESTIMATE FOR ALL	ACTUAL REPETITION, PUN. CONN. 5 OR > REMOVED	EXCESS OF ACTUAL OVER EXPECTATION BY CHANCE AND NUMBER FAVORITISM
	SET A $n = 245$	SET B $n = 245$	SET C $n = 245$	SET D $n = 245$			
Ah	128	120	117	192	140	249	109
An	119	119	117	121	120	151	31
Bi	104	116	143	113	120	170	50
Br	126	163	157	185	160	203	43
Br 2	151	130	131	133	135	241	106
Bro	111	112	106	112	110	187	77
Burg	162	143	166	168	160	428	268
Burl	129	116	119	121	120	202	82
Burk	169	135	116	169	147	329	182
Ca	108	120	116	109	115	131	16
Co	137	122	146	148	135	236	101
F	139	134	139	122	135	229	94
G	151	119	161	136	142	315	173
H	109	113	134	114	118	150	32
Le	129	112	124	108	118	161	43
Lip	120	121	135	123	125	259	134
Ma	118	146	145	138	137	272	135
Me	203	121	141	121	147	315	168
N	121	126	134	122	126	187	61
O	385	281	122	151	235	372	137

with some memory or other equivalent of the situation, the punishment has zero influence upon the punished connection.

If this is so, the strengthening observed in these experiments is a measure of the strengthening due to a mere occurrence, provided proper allowance is made for (*a*) spread or scatter beyond four steps, (*b*) the possible persistent suggestion of a particular response (here a number, a part of a diagram or a position in a line) by the situation, and (*c*) a possible occasional confirming reaction due to the satisfyingness of making any choice and response.

Observation and experiment can, with enough labor, learn what allowances, if any, are needed for (*a*) and (*b*). We have evidence already that there is no spread or scatter of the confirming influence of a reward beyond four steps in many of the seventeen relevant experiments in which there is certainly a large excess over chance, and almost certainly an excess over chance plus number favoritism. There is also evidence that some persons who are entirely unaware of any persistent suggestions of particular responses of the sort described in (*b*) show large excesses over chance plus number favoritism. Though doubtless a real cause, *b* does not seem likely to be an adequate one.

I have not found it possible, within such multiple-choice experiments as these, to distinguish between a confirming reaction due to the satisfyingness of making *any* choice and response and a strengthening influence due to the mere use of a connection.

The former is conceivable; it would permit the explanation of all strengthening of connections by one physiological function or mechanism; it would explain the phylogenetic development of learning by one variation, the appearance of a confirming reaction, rather than by two; it would avoid the difficulty of explaining why the mere use of some connections does strengthen them whereas the mere use of others does not. But there is no direct evidence for it in the experiments. The subjects feel no observable satisfaction at saying *any* number or marking *any* part of a diagram. Their satisfaction seems to them limited to the connections that are rewarded (or that, in later stages of an experiment, they know will be rewarded, even before they make them). And, in general, the evidence here seems to be against it, as in the experiments of Chapter IV of the *Fundamentals of Learning*, and as in the formation of conditional reflexes by Cason ['22], by Hudgins ['33] and by others who have used techniques in which the subjects gain nothing by the formation of such connections.

There is still another possibility allied to (*c*). The mere occurrence of a connection may strengthen it and yet do so indirectly and occasionally by arousing the confirming reaction, not directly and uniformly by the action of the connection itself. I have elsewhere suggested, as a working hypothesis to explain the "spread" or "scatter" phenomenon, that the confirming reaction, being a physiological rather than a logical influence, occasionally misses the connection which the reward follows and belongs to, and influences by mistake a punished connection coming a little earlier or later. In a somewhat similar way the confirming reaction itself, which is in such experiments probably in a state of extreme excitability or sensitiveness or readiness to act, may occasionally be set off, not by its normally adequate stimulus, a state of affairs satisfying to the mental set or purpose or controlling powers, but by mistake by stimuli that normally would be inadequate. This hypothesis would have the merits noted above for (*c*) and would perhaps be consistent with the facts of other experiments. But discussion of all these possibilities may well be deferred until further experiments bearing upon them have been made.

[1] Some of these are reported in *An Experimental Study of Rewards* by Thorndike and others, 1933; others have been or will be published in various journals.

[2] Some of this strengthening may be due to repetition by the subjects of the numbers after the announcements of right. There was little time for them to do this, and there is no surety that those who did it profited by it, but it was not forbidden and was often attempted. For the purposes of this article, we do not care just what caused the 43.6 per cent of repetitions of a rewarded occurrence.

[3] For 16 of the 24 subjects, every one of the 4800 responses was tabulated. For the

other 8 subjects enough were tabulated (800 for two, 1200 for three, 2400 for one and 3800 for two) to satisfy all requirements. In all cases, the first two sets of 200 each and the last two sets were tabulated. In the case of the two for whom only 800 were tabulated, there was close agreement among the four sets of 200.

[4] The complete detailed records are kept on file in the office of the Division of Psychology of the Institute of Educational Research at Teachers College for use by any one who wishes to make independent estimates.

[5] First occurrences only are used; and, of course, the count of repetitions is for the trial or round following the occurrence.

[6] Experiments J, L and M are described in *An Experimental Study of Rewards*.

[7] In a study as yet unpublished.

LITERATURE CITED

Cason, H., "Conditioned Pupillary Reactions," *J. Exper. Psychol.*, **5**, 108–146 (1922).

Hudgins, C. V., "Conditioning and the Voluntary Control of the Pupillary Light Reflexes," *J. Gen. Psychol.*, **8**, 3–51 (1933).

Lorge, I., "The Efficacy of Intensified Reward and of Intensified Punishment," *J. Exp. Psych.*, **16**, 177–207 (1933).

Lorge, I., and Thorndike, E. L., "The Comparative Strengthening of a Connection by One or More Occurrences of It in Cases Where the Connection Was Punished and Was neither Rewarded nor Punished," *J. Exp. Psych.*, **16**, 374–382 (1933).

Thorndike, E. L., and the Staff of the Division of Psychology, of the Institute of Educational Research, "The Fundamentals of Learning" (1932).

Thorndike, E. L., "An Experimental Study of Rewards," Teachers College, Columbia University Contributions to Education, No. 580 (1933).

THE EFFECTS OF THE PITUITARY HORMONES ON THE MELANOPHORES OF FISHES

BY J. M. ODIORNE

ZOÖLOGICAL LABORATORIES, HARVARD UNIVERSITY

Communicated May 27, 1933

In contrast to the uniform effect which adrenalin has on the melanophores of many animals, the results obtained by the use of extracts from the pituitary body have frequently been contradictory. In view of this fact it seems appropriate to publish an account of experiments made within the last three years which put the situation in a somewhat different light. Most of the experiments have been carried out on Fundulus heteroclitus and Ameiurus nebulosus, but occasionally other fishes have been tested. I am pleased to acknowledge the kind suggestions made by Professor G. H. Parker, under whose direction the work has been done.

Spaeth (1918) reported that the melanophores on isolated scales of Fundulus contracted when placed in a pituitary solution. This was confirmed by Wyman (1924), who found that injection of the solution into the body cavity brought about a paling of the fish; similar results were ob-

tained by Gilson (1926). Nevertheless, all experiments which I have made have been negative; the melanophores of isolated scales failed to contract in post-pituitary solutions and fishes injected with such solutions showed no pigmentary response.

"Pituitrin," an aqueous solution containing the water-soluble principles of the posterior lobe of the pituitary gland, was the first extract used. If scales from a white-adapted Fundulus are placed in distilled water the melanophores remain contracted, while the melanophores of scales placed in Ringer's solution expand within a few minutes. Consequently, if pituitrin is added to the Ringer's solution one would expect the melanophores to contract. Spaeth reported that they contracted completely and that the time required for this response was a function of the concentration of the pituitary solution. In spite of this, in my experiments such a result was not obtained even when equal volumes of pituitrin and Ringer's solution were used. Yet if a small amount of adrenalin was added to such a solution nearly all of the melanophores contracted within five minutes. When pituitrin was added to distilled water in which scales had been placed the melanophores remained unaltered, showing that it had no expanding effect. The melanophores of scales transferred from distilled water to Ringer's solution expanded, which shows that the melanophores were not seriously harmed by distilled water.

The results of experiments on isolated scales were confirmed by intraperitoneal and intramuscular injections of pituitrin. Dark fishes from a black dish injected with pituitrin (in doses of from 0.1 to 0.4 cc.) and returned to the black dish remained dark. If transferred to a white dish they rapidly became light. Conversely, light fishes which were injected with pituitrin remained light if returned to a white dish, but rapidly darkened when transferred to a black dish. In other words, they behaved like normal uninjected fishes, and when in the same dish with controls could not be distinguished from them by their color.

That the pituitrin was capable of producing pigmentary responses in suitable animals was demonstrated by experiments conducted simultaneously on Ameiurus. The responses of these fishes to backgrounds are slow as compared to those of Fundulus, but they become light greenish on a white background and extremely black on a black background. Light fishes injected with pituitrin became darker, the degree varying with the size of the dose (usually 0.2 cc. produces maximal darkening of a fish six to eight inches long; the same effect can be produced in smaller fishes by 0.1 cc.). The response was slow, requiring an hour or more for its completion, but the fishes remained dark for hours after the injection. As a check on experiments with Fundulus, pituitrin from the same ampoule was injected into Ameiurus with the result described, so that the efficacy of the preparations used in these tests was certain.

In addition to pituitrin, preparations of its two principles are available. "Pitressin," a solution containing the pressor principle (beta-hypophamine), formerly known as vasopressin, produced results in Ameiurus like those brought about by pituitrin, and failed to produce any effect on Fundulus. "Pitocin," formerly called "oxytocin," represents the oxytocic principle (alpha-hypophamine); it produced no effect on Ameiurus, and was not tried on Fundulus.

Experiments on eels and black-backed goldfishes gave results like those obtained on Ameiurus. The pigmentary responses of eels are even slower than those of Ameiurus, though goldfishes react at about the same rate.

An extract of the anterior lobe of the pituitary gland is also available, and the effects of this preparation, which is known as "antuitrin," have also been investigated with interesting results. In Fundulus the injection of 0.2 cc. of antuitrin produced contraction of the melanophores of dark specimens kept on a black background. The response was prompt but not all of the melanophores were concerned in it, for the fishes presented a spotty appearance unlike the even light color which results from the injection of adrenalin. The pale condition did not persist for more than two hours.

Experiments using isolated scales furnished confirmation. When antuitrin was added to a small amount of Ringer's solution in which scales had been placed the expanded melanophores contracted within a few minutes. Thus it gave results similar to those obtained by Spaeth with pituitrin.

It seems likely that the dissimilarity between the results obtained by Spaeth and Wyman and those described here can be traced to differences in the preparations used. Spaeth (1918) employed "pituitrin" (Parke, Davis and Co.) and a pituitary extract furnished by H. K. Mulford & Co.; as the title of his paper indicates, he believed that the melanophores of Fundulus could be used for the biological standardization of pituitary extracts and other drugs. The samples which he used had been standardized by other methods, yet he wrote that "commercial samples of pituitary extract which were supposedly of the same physiological strength were found, by the use of the melanophore method, to vary widely." If we assume that these early preparations contained not only the principles of the posterior lobe but substances from the anterior lobe as well, then Spaeth's results can be explained and we can understand why supposedly standardized samples should "vary widely." Instead of checking the strength of the extracts he was measuring their impurity with regard to material from the anterior lobe. Such an assumption is not unwarranted in view of the anatomical relationships of the various parts of the pituitary and the diffusion of substances from one part to another which has been amply demonstrated. Wyman (1924) in his experiments used pituitary

preparations which were supplied to him by Spaeth and which were from the same lot used by Spaeth for his 1918 paper, so that it is not surprising that their results were similar.

The first report that pituitary extract causes expansion of the melanophores of fishes was published by Abolin (1925). Following injections of "infundin" he observed expansion of these cells in Phoxinus, though the response was modified by the activity of other chromatophores. He obtained similar results using Nemachilus barbatula, Esox lucius, Carassius vulgaris and Leuciscus rutilis. Hewer (1926) found, on the contrary, that extracts of the posterior lobe of the cod pituitary produced contraction of the melanophores in Phoxinus. Infundin gave rather inconclusive results. Giersberg (1930) also studied Phoxinus and reported that infundin produces the effect claimed by Abolin; in the perch no such positive effect was obtained. In Gobius and Pleuronectes pituitrin and "hypophysin" cause contraction of the melanophores, according to Meyer (1931). Lundstrom and Bard (1932), working on the dogfish Mustelis canis, added the elasmobranchs to the list of animals which respond to pituitary substances. The pale condition resulting from extirpation of the entire hypophysis or of the posterior (neuro-intermediate) lobe is replaced by the normal dark color following injections of suspensions of the dogfish posterior lobe, or of infundin, pituitrin and pitressin, but pitocin has no effect.

From this we can at least conclude that different kinds of fishes respond in very different ways to pituitary preparations. Of the various fishes mentioned some do not respond to pituitrin at all, in the majority the melanophores expand and the fish becomes darker, in others (Gobius and Pleuronectes) the melanophores contract, and in the case of Phoxinus the effect is disputed. In Fundulus the result previously attributed to pituitrin seems to be due to some principle contained in antuitrin. Recently it has been estimated that twenty-two pituitary hormones have been reported by various investigators. In view of this it does not seem surprising that more uniform results have not been obtained. Very likely some fishes respond to one principle but not to another, and it is not certain that all fishes would respond in the same way to a single substance.

Hogben (1924), in his monograph "The Pigmentary Effector System," wrote as follows: "One other fact emerging from Spaeth's work (1917) is of special interest, because it shows very clearly that the physiological properties of Fish and Amphibian melanophores are not to be regarded as identical. In the case of Amphibia, we have seen that pituitary extract exerts a specific and very characteristic action in promoting melanophore expansion. Spaeth, who has suggested a method of pharmacological assay based upon his observations, claims that pituitary extract induces contraction of the melanophores in fish scales. While there is reason to believe that the melanophores of Fishes and Amphibia respond in a different

manner to other stimuli, this appears to be a specially clear-cut instance of dissimilarity, since there is satisfactory evidence that the action of pituitary extract upon Amphibian melanophores is local." Later he concludes: "If Spaeth's observations on the pituitary reaction of Fish melanophores are correct, the mechanism of pigmentary response in Amphibia and Fishes must differ profoundly."

In view of the results obtained on various fishes since that time it seems clear that, although the melanophores of fishes and amphibians are perhaps not to be regarded as identical, they are not profoundly different. Indeed, the unanimity of response to pituitary extracts in the amphibians has been set aside by the work of Collins and Adolph (1926), who reported that the melanophores of the salamander Diemyctylus contract under the influence of pituitrin. Consequently the most that can be said is that not all fishes and amphibians respond to pituitary substances in the same way.

Even though the difference between amphibians and fishes is not as great as was previously thought, it is nevertheless true that the amphibians as a group are more uniform than the fishes in this respect. In Fundulus the pigmentary responses have been shown to be to a great extent under the control of the nervous system. Though the nervous system also plays a part in the control of the color changes of Ameiurus and the eel, the slower responses of these forms suggest the possibility that here the nervous system is materially aided by hormones. Possibly the differences in the responses of fishes to pituitary substances are somehow related to the varying importance of the nervous system and the humoral system in these forms.

Abolin, L., "Beeinflussung des Fischfarbwechsels durch Chemikalien. I. Infundin- und Adrenalinwirkung auf die Melano- und Xanthophoren der Elritze (Phoxinus laevis Ag.)," *Arch. mikr. Anat. Entw.-mech.*, **104**, 667–698 (1925).

Collins, H. H., and Adolph, E. F., "The Regulation of Skin-Pattern in an Amphibian, Diemyctylus," *Jour. Morph.*, **42**, 473–522 (1926).

Giersberg, H., "Der Farbwechsel der Fische," *Zeit. vergl. Physiol.*, **13**, 258–279 (1930).

Gilson, A. S., Jr., "The Control of Melanophore Activity in Fundulus," *Jour. Exp. Zoöl.*, **45**, 457–468 (1926).

Hewer, H. R., "Studies in Colour Changes of Fish. I. The Action of Certain Endocrine Secretions in the Minnow," *Brit. Jour. Exp. Biol.*, **3**, 123–140 (1926).

Hogben, L. T., *The Pigmentary Effector System*, Edinburgh, 152 pp. (1924).

Lundstrom, H. M., and Bard, P., "Hypophysial Control of Cutaneous Pigmentation in an Elasmobranch Fish," *Biol. Bull.*, **62**, 1–9 (1932).

Meyer, E., "Versuche über den Farbwechsel von Gobius und Pleuronectes," *Zool. Jahrb., Abt. Allgem.*, **49**, 231–270 (1931).

Spaeth, R. A., "Concerning a New Method for the Biological Standardization of Pituitary Extract and Other Drugs," *Jour. Pharm. Exp. Therap.*, **11**, 209–219 (1918).

Wyman, L. C., "Blood and Nerve as Controlling Agents in the Movements of Melanophores," *Jour. Exp. Zoöl.*, **39**, 73–132 (1924).

THE OCCURRENCE OF GUANOPHORES IN FUNDULUS

BY J. M. ODIORNE

ZOÖLOGICAL LABORATORIES, HARVARD UNIVERSITY

Communicated May 27, 1933

Though the changes in shade and color which occur in Fundulus heteroclitus in response to backgrounds and other stimuli have hitherto been accounted for by changes in the melanophores and xanthophores, this fish also possesses chromatophores which apparently have been neglected if not entirely overlooked. Perhaps this is because of the distribution of these cells and their position in relation to the melanophores. The object of this report is to describe briefly some of their morphological and physiological characteristics, and to point out their significance in bringing about alterations in the appearance of the fish.

If one of the large scales in the mid-dorsal region just behind the head is removed from an individual adapted to a black background, and the area thus exposed is observed with a microscope by reflected light, it will be seen to present the features illustrated by figure 1. The most conspicuous structures are the expanded melanophores; their processes partly conceal the xanthophores which lie among them, but deeper in the dermis. Imbedded in many of the melanophores are the iridosomes which have been the subject of a recent paper by Foster (1933). According to the terminology of Ballowitz (1912) an iridosome is a group of "iridocytes," and when associated with a melanophore the combination becomes a "melaniridosome."

If instead of a black-adapted specimen a fish from a white background is examined in the same way, a much different condition is revealed (Fig. 2). The melanophores are contracted and appear to be surrounded by silvery halos, which are actually the radiating processes of chromatophores lying beneath the melanophores. These processes form an interlacing network in many places and hide the deeper xanthophores from view.

If a piece of integument is removed and studied by transmitted light, the chromatophores, which always appear silvery (or luminously white) by reflected light, lose their brilliance and instead are a dull yellowish gray. They are never iridescent, though their brilliant whiteness is due to their ability to reflect incident light. If examined under oil immersion the substance which they contain is seen to be in the form of small particles. These may appear to be closely packed together or they may show Brownian movement. In either case it is difficult to determine whether they are merely granular or definitely crystalline, for they are smaller than the melanin granules. Some writers speak of amorphous guanin, while others

hold the opinion that it is always crystalline but, whatever its state, it seems likely that the substance is guanin.

Consequently I shall refer to them as "guanophores," a term used by Schmidt (1912) to indicate chromatophores (in amphibians) containing guanin. Ballowitz (1912) applied this term to similar chromatophores of fishes which had been called "iridocytes" by Pouchet (1876), but Ballowitz continued to use Pouchet's term in his numerous papers. Both Connolly (1925) and Fries (1931) have referred to the iridosomes of Fundulus as "guanophores or iridocytes," but in spite of the fact that an iridocyte is a type of guanophore it seems desirable to use the word "iridocyte" only when referring to a guanophore which is iridescent. Biedermann (1926) has already pointed out that there are guanophores which show no interference colors whatever, and the guanophores of Fundulus which are the subject of this paper are of this type. Gilson (1926), in describing the iridocytes of Fundulus, said: "By reflected light the iridocytes appear shining and silvery, but by transmitted light they are brownish in appearance and, in a hasty examination of tissue, may even be mistaken for melanophores." He did not describe their location, neither did he give illustrations, but from his description it would seem that he observed these structures though he did not recognize them as different from the true iridocytes.

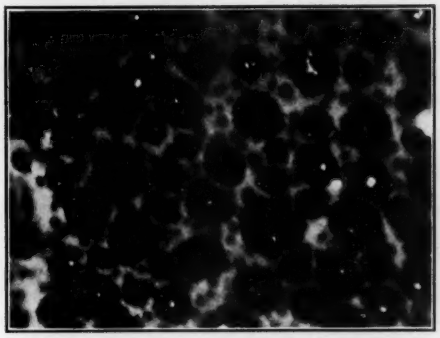


FIGURE 1

Photomicrograph of the area exposed by the removal of a scale in the mid-dorsal line of a black-adapted Fundulus. The white spots in many of the melanophores are the iridosomes. The xanthophores appear as small, dark gray bodies.

Ginsburg (1929) has described the development of the chromatophores in the guppy, Lebistes reticulatus, and he has recognized two types of "guaninophores," as he called the guanin-containing cells. In addition to those which show interference colors there are others which are generally much larger, and which appear dull white or silvery white by incident light. Recently Sumner and Wells (1933) have described and illustrated structures in Lebistes which they have interpreted as "iridosomes," following the terminology of Ballowitz. Apparently, they are the silvery white "guaninophores" mentioned by Ginsburg. The guanophores of

Fundulus bear a very close resemblance to these "iridosomes" or "guaninophores" of Lebistes.

Ballowitz, in a series of contributions, has described chromatophores and chromatophore combinations in a number of different fishes, and an examination of these papers makes it clear that iridosomes may not only be dissimilar in different genera, but also that a single fish may have more than one type, as in Serranus. The melaniridosomes of Fundulus most closely resemble those of Perca and Acerina (Ballowitz, 1914, Taf. I. Figs. 1–5). The guanophores of Fundulus are somewhat similar to the iridosomes of Gobius (Ballowitz, 1913, Taf. XI, Figs. 72, 73 and Taf. XII, Figs. 94a, 94b); to be sure the optical properties are not the same, but the configuration and position are comparable and both expand and contract.[1] Ballowitz did not investigate the effects of different backgrounds on the iridosomes of Gobius, but Meyer (1931) reported that they expanded in response to a white background and contracted under the influence of a black one.

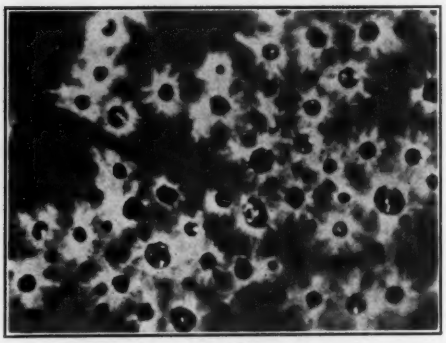


FIGURE 2

Photomicrograph of an area similar to that shown in figure 1, after the application of adrenalin. The melanophores are contracted and the guanophores are widely expanded, as in a fish adapted to a white background.

The guanophores of Fundulus are most numerous on the dorsal surface though they extend downward over the sides for some distance. In females they are absent or rare in the fins and tail but in males small guanophores, usually without overlying melanophores, are found in the dorsal, anal and caudal fins. They occur in areas which are free from melanophores (Fig. 3), frequently accompanied by iridescent patches, and intensify the color pattern which is primarily caused by the absence of the melanophores. Isolated guanophores are occasionally found on the body as well. In either situation they are small and when expanded do not exhibit such extensive processes as do those below melanophores. In some fishes guanophores also occur below the melanophores which lie in the dermis external to the scales, and near their edges. If a scale bearing such guanophores is removed and examined from the under side, they will be found to be small. They are located near the margin of the pigment mass (in a contracted melanophore) and usually do not cover its entire under surface.

The guanophores seem to be most numerous where the stratum argenteum is poorly developed, though exception must be made of the fins, especially those of the females. Very likely the reason that they have escaped notice is that they are rare or absent in those parts of the body which lend themselves to the study of melanophores and xanthophores, namely, the isolated scales and the caudal fin. Besides they are not conspicuous when viewed by transmitted light.

That the guanophores of Fundulus do have the ability to expand and contract is shown by the following experiment. If a few drops of adrenalin (1:1000) are put on an area such as that illustrated in figure 1 the migration of the melanin granules toward the centers of the melanophores can be plainly seen within a short time. Presently the outward migration of the guanin can be discerned and within about five minutes the appearance of the area is similar to that shown in figure 2. From this it is clear that these structures are chromatophores and are not merely immobile deposits of guanin beneath the melanophores which are exposed and concealed by the movements of the pigment within these cells. The isolated guanophores are favorable for observing these movements of the guanin, because of the absence of overlying melanophores, but as already pointed out the processes are rather small.

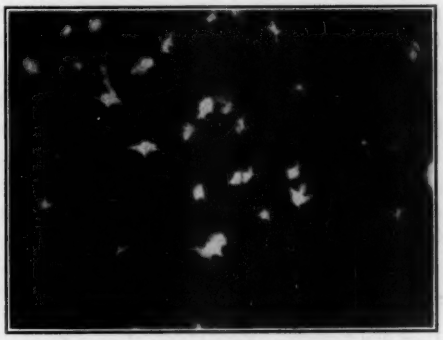


FIGURE 3

Photomicrograph of a portion of the dorsal fin of a male Fundulus, showing isolated guanophores.

The experiment just described also demonstrates that the guanophores respond to adrenalin, but the response is opposite to that of the melanophores. The same reaction occurs if 0.2 cc. of adrenalin (1:1000) is injected into the body of a dark fish, instead of being applied to the naked dermis.

"Antuitrin," an extract of the anterior lobe of the pituitary body, causes contraction of the melanophores of Fundulus when injected into the body cavity or when applied to the naked dermis (see paper preceding this). The response is not as thorough as the response to adrenalin but is perfectly definite. The guanophores expand under these conditions.

In a fish from a yellow background the melanophores are contracted though not always maximally; at the same time the guanophores are

partially expanded but are only slightly visible. Consequently the xanthophores are not concealed by either melanophores or guanophores and the animal assumes a decidedly yellowish green color.

For the detailed study of the guanophores it is necessary to remove the overlying scales but this does not mean that their presence is not evident without this procedure. Different individuals vary with respect to the completeness with which they adapt their color to a white background, and with respect to the time required to become maximally light, but the guanophores are certainly important factors in this adaptation, even though not all fishes assume the decided grayish color frequently seen. Situated as they are below the melanophores, the guanophores do not detract from the dark appearance of fishes in which the melanophores are expanded; in light-adapted fishes the greatly expanded guanophores increase the paleness caused by contraction of the melanophores. Furthermore, the reactions of the guanophores to various stimuli have so far been found to be just opposite to those of the melanophores, a fact which shows clearly that the chromatophores of this fish are far from being physiologically alike and that their behavior is decidedly "adaptive."

[1] It has become customary to refer to the changes which take place in chromatophores as "expansion" and "contraction," though this does not imply that the shape is actually altered. Instead, the pigment within the cell shifts, now being massed in the center (contraction), now distributed in the processes (expansion).

Ballowitz, E., "Über chromatische Organe in der Haut von Knochenfischen," *Anat. Anz.*, **42**, 186–190 (1912).

Ballowitz, E., "Über schwarzrote und sternförmige Farbzellenkombinationen in der Haut von Gobiiden," *Zeitsch. wiss. Zool.*, **106**, 527–593 (1913).

Ballowitz, E., "Die chromatischen Organe, Melaniridosomen, in der Haut der Barsche (Perca und Acerina)," *Ibid.*, **110**, 1–35 (1914).

Biedermann, W., "Vergleichende Physiologie des Integuments der Wirbeltiere. Zweiter Teil. Die Hautfärbung der Fische, Amphibien und Reptilien," *Ergeb. der Biol.*, **1**, 174–342 (1926).

Connolly, C. J., "Adaptive Changes in Shades and Colour of Fundulus," *Biol. Bull.*, **48**, 56–77 (1925).

Foster, K. W., "Color Changes in Fundulus with Special Reference to the Color Changes of the Iridosomes," *Proc. Nat. Acad. Sci.*, **19**, 535–540 (1933).

Fries, E. F. B., "Color Changes in Fundulus, with Special Consideration of the Xanthophores," *Jour. Exp. Zoöl.*, **60**, 389–426 (1931).

Gilson, A. S., Jr., "Melanophores in Developing and Adult Fundulus," *Jour. Exp. Zoöl.*, **45**, 415–455 (1926).

Ginsburg, J., "Beiträge zur Kenntnis der Guaninophoren und Melanophoren, *Zool. Jahrb., Abt. Anat.*, **51**, 227–260 (1929).

Meyer, E., "Versuche über den Farbwechsel von Gobius und Pleuronectes," *Zool. Jahrb., Abt. Allgem.*, **49**, 231–270 (1931).

Pouchet, G., "Des changements de coloration sous l'influence des nerfs," *Jour. Anat. Physiol.*, **12**, 1–90, 113–165 (1876).

Sumner, F. B., and Wells, N. A., "The Effects of Optic Stimuli upon the Formation and Destruction of Melanin Pigment in Fishes," *Jour. Exp. Zoöl.*, **64**, 377–394 (1933).

PHOTIC STIMULATION AND RHYTHMICAL CONTRACTIONS OF THE MANTLE FLAPS OF A LAMELLIBRANCH

BY JOHN H. WELSH

ZOÖLOGICAL LABORATORIES, HARVARD UNIVERSITY

Communicated June 2, 1933

Ortmann (1911) describes a peculiar rhythmical movement of the mantle flaps of females of two members of the genus *Lampsilis* (*L. ventricosa* and *L. multriradiata*), fresh-water lamellibranchs. This movement was noted by the author in a female of *Lampsilis nasuta*. With the knowledge that the siphons, mantle edge and labial palps of bivalve molluscs often contain photoreceptors, it occurred to the writer to determine whether or not this rhythmical movement was influenced by light intensity. Such movement was being sought for in order to observe the effect of light intensity on frequency of muscular contractions where amplitude was not involved as it is in leg movement (Welsh, 1932). A few preliminary trials indicated that light did play an important rôle in determining the frequency of these rhythmical contractions. A series of determinations of the time required for ten movements, over a fairly wide range of intensities, is contained in table 1. The frequency of movement is shown plotted

TABLE 1

TIMES IN SECONDS FOR TEN MOVEMENTS OF THE MANTLE FLAPS OF LAMPSILIS WITH THEIR AVERAGES, AND THE FREQUENCY (NUMBER OF MOVEMENTS PER SECOND) AT EACH OF SEVERAL INTENSITIES OF ILLUMINATION. TEMP. 21.3°C.

INTENSITY (FOOT-CANDLES)	0.20	0.41	0.83	1.66	3.6	12.0	23.0	180.0	689.0
Time (secs.) for 10 movements	65.0	48.0	28.6	24.5	22.5	21.2	21.0	20.1	20.0
	61.0	46.8	30.5	25.8	22.3	20.9	20.8	20.1	20.0
	62.0	49.2	32.2	27.4	22.2	21.4	20.8	20.1	20.2
	63.5	48.6	35.2	26.6	22.1	21.5	20.6	20.1	20.4
	61.8	46.7	33.0	26.4	21.9	21.8	20.9	20.5	20.1
			32.5	27.6	22.6	21.5	20.9	20.2	19.7
			30.8	27.6	23.4	21.2	21.0	20.4	19.6
			35.0	27.4	23.6	21.5	20.6	19.9	19.8
			36.0	27.0	23.3	21.0	20.6	20.4	20.0
			38.0	25.6	23.6	21.2	21.2	20.1	19.8
Averages	62.66	47.86	33.18	26.59	22.75	21.32	20.84	20.19	19.96
Frequency	0.159	0.210	0.301	0.376	0.439	0.469	0.481	0.495	0.502

against the logarithm of the light intensity in figure 1. The curve is similar to that obtained in a study of frequency of leg movement, as a function of light intensity, in a water mite *Unionicola* (Welsh, 1932), except that the lower portion is incomplete as the rhythm was interrupted at low light

intensities, and ceased entirely after a short exposure to an illumination of about 0.2 foot-candles. Five consecutive readings at the two lower intensities were as many as could be obtained without interruption.

It was impossible to carry out check experiments on other individuals because of their rareness locally, and the fact that the contractions occur only for a period of a few weeks when the marsupia are loaded with developed glochidia and these are being discharged.

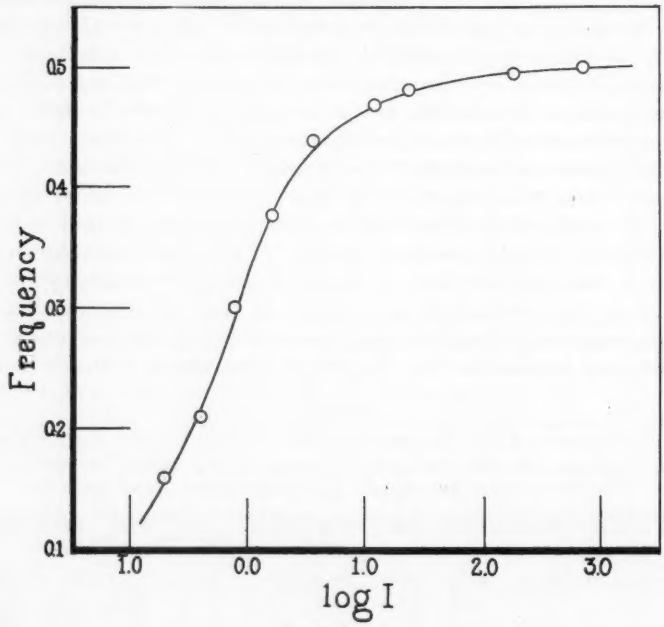


FIGURE 1

Plot of data of table 1 showing frequency of movement (number of movements per second) of the mantle flaps of *Lampsilis nasuta*, plotted against the logarithm of the light intensity. Observations were begun at the highest illumination.

Ortmann (1911) believed this movement to be for purposes of aeration of the glochidia during the breeding season. This seems to be the most logical explanation as the action of the mantle flaps greatly increases the amount of water passing through the mantle cavity; however, if this is its only function it is difficult to understand why it stops at low intensities of illumination or in darkness. In nature the movements must occur only during daylight hours, and during the night the circulation must be the result solely of ciliary activity. The phenomenon deserves further in-

vestigation as other factors such as CO_2 tension would doubtless be found to influence the frequency of movement as Redfield (1917) found in a study of rhythmic contractions in the mantle of certain lamellibranchs, a quite different process but one having a similar end result.

The very precise regularity of the movement nearly equals the regularity of heart beat and provides a very convenient rhythm for use in an analysis of a number of phases of the physiological environment of such fresh-water organisms. It is proposed to investigate further this rhythmical movement when more breeding females are obtained.

Ortmann, A. E., "Monograph of the Najades of Pennsylvania, Parts I and II," *Memoirs Carnegie Museum*, **4**, 279–347 (1911).

Redfield, E. S. P., "The Rhythmic Contractions in the Mantle of Lamellibranchs," *Jour. Exp. Zoöl.*, **22**, 231–239 (1917).

Welsh, J. H., "Photokinesis and Tonic Effect of Light in Unionicola," *Jour. Gen. Physiol.*, **16**, 349–355 (1932).